CANADIAN JOURNAL OF MATHEMATICS

Journal Canadien de Mathématiques

VOL. VIII - NO. 2
1956

Published for

THE CANADIAN MATHEMATICAL CONGRESS

by the

University of Toronto Press

The chief languages of the *Journal* are English and French.

Manuscripts for publication in the *Journal* should be sent to the *Editor-in-Chief*, H. S. M. Coxeter, University of Toronto. Everything possible should be done to lighten the task of the reader; the notation and reference system should be carefully thought out. Every paper should contain an introduction summarizing the results as far as possible in such a way as to be understood by the non-expert.

All other correspondence should be addressed to the *Managing Editor*, G. de B. Robinson, University of Toronto.

The *Journal* is published quarterly. Subscriptions should be sent to the *Managing Editor*. The price per volume of four numbers is $8.00. This is reduced to $4.00 for individual members of recognized Mathematical Societies.

The Canadian Mathematical Congress gratefully acknowledges the assistance of the following towards the cost of publishing this Journal:

University of Alberta

Assumption College
Carleton College
Université Laval
University of Manitoba
McMaster University
Queen's University
St. Mary's University
University of British Columbia
Ecole Polytechnique
Loyola College
McGill University
Université de Montréal
Royal Military College
University of Toronto

National Research Council of Canada
and the
American Mathematical Society

AUTHORIZED AS SECOND CLASS MAIL, POST OFFICE DEPARTMENT, OTTAWA

he
ng
on
er
le

ng

nt
rs
of

he
l:

a

ON CONNECTIONS OF CARTAN

SHÔSHICHI KOBAYASHI

Introduction. Consider a differentiable manifold M and the tangent bundle $T(M)$ over M, the structure group of which is usually the general linear group G'. Let P' be the principal fibre bundle associated with $T(M)$. Consider the fibre F of $T(M)$ as an affine space, then we have acting on F the affine transformation group G, which contains G' as the isotropic subgroup. Following the idea of Klein, it is more natural to take G as the structure group of the bundle $T(M)$. Let P be the principal fibre bundle associated to $T(M)$ with group G.

In the classical theory of affine connections, there are two points of view. The one is due to Levi-Civita, who considered each tangent space of M as a vector space and explained a connection as a law of parallel displacement of vectors along curves. From the point of view of the theory of connections in fibre bundles, a connection in the sense of Levi-Civita is a connection in the principal fibre bundle P' with group G'. The other point of view is due to E. Cartan. Following him, each tangent space of M is an affine space on which the affine transformation group G acts transitively, and an affine connection is a law of development of tangent spaces along curves; it is a connection in P.

The idea of Cartan was rigorously established by Ehresmann (**3**) as follows. Consider a fibre bundle B satisfying the conditions of *soudure* (see §2); the fibre F is homeomorphic to a homogeneous space G/G' and the structure group G of B can be reduced to G'. As in the case of tangent bundle, we obtain two principal fibre bundles P and P' with group G and G' respectively and P' is contained in P. A connection in P is called a connection of Cartan, if it satisfies the following condition: the differential form ω defining the connection gives an absolute parallelism on P'. The importance of this condition was shown in previous papers (**4**; **5**).

It is known that there is a correspondence between affine connections in the sense of Cartan and those in the sense of Levi-Civita; there is a canonical one-to-one correspondence between the set of connections in P and the set of connections in P' (**7**).

The purpose of the present paper is to show that there exists a one-to-one correspondence between the set of Cartan connections in P and the set of infinitesimal connections in P', if the homogeneous space $F = G/G'$ is *weakly reductive* (see §2). We shall show also that in such a case the torsion forms can be defined. The last section will be devoted to the application to invariant connections.

Received April 25, 1955.

1. Tangent vectors. The manifolds and the mappings considered in this paper are all of class C^∞. For the definition of tangent vector and the differential of a mapping, the reader is referred to Chevalley's book **(2)**.

Let M be a manifold. We denote by $T(M)$ the set of all tangent vectors to M. For any two manifolds M and M', we have a natural isomorphism

$$T(M \times M') = T(M) \times T(M').$$

Let G be a Lie group and $\phi: G \times G \to G$ be the mapping defining group operation:

$$\phi(s, s') = s \cdot s', \qquad s, s' \in G.$$

Consider the differential mapping[1] $\delta\phi: T(G \times G) \to T(G)$. $T(G \times G)$ being identified with $T(G) \times T(G)$, $\delta\phi$ can be considered as a mapping of $T(G) \times T(G)$ onto $T(G)$ and defines a group operation in $T(G)$. The Lie group $T(G)$, obtained in this way, is called the *tangent group to* G. We have a natural imbedding of G into $T(G)$ and G is considered as a subgroup of $T(G)$. The set of all tangent vectors to G at the unit, which we shall denote by $T_e(G)$, is a normal subgroup of $T(G)$ and will be identified with the Lie algebra of G.

Suppose G acts, as a transformation group, on a manifold P on the right and let $\psi: P \times G \to P$ be the mapping defining the transformation law. Then, the differential mapping

$$\delta\psi: T(P) \times T(G) \to T(P)$$

defines $T(G)$ as a transformation group on $T(P)$ acting on the right. If P is a principal fibre bundle over M with group G and with projection π, then $T(P)$ is a principal fibre bundle over $T(M)$ with group $T(G)$ and with projection $\delta\pi$.

2. Soudure. Let B be a fibre bundle over base manifold M, with fibre F and with Lie structure group G. B is *soudé* **(3)** to M, if the following conditions are satisfied:

(s.1) G acts on F transitively: then F can be identified with the homogeneous space G/G', where G' is the isotropic group at a point o of F.

(s.2) $\dim F = \dim M$.

(s.3) The structure group G of the bundle B can be reduced to G': in other words, B admits a cross-section, which we shall denote by σ. When B is considered as the fibre bundle with structure group G', it will be denoted by B'.

(s.4) Two fibre bundles $T(M)$ and $T_F(B)$ over M, with group $GL(n, R)$ (where $n = \dim M$), are equivalent, where $T(M)$ is the space of all tangent vectors to M and $T_F(B)$ the space of all tangent vectors to F_x at $\sigma(x)$, x running through M.

Let P (resp. P') be the principal fibre bundle associated to B (resp. B').

[1]Chevalley denotes the differential of ϕ by $d\phi$.

The structure group and the fibre of P (resp. P') are G (resp. G'). P' can be considered as a submanifold of P.

Let $\mathfrak{g}$, $\mathfrak{g}'$ be the Lie algebras of G and G' respectively. Take a vector subspace $\mathfrak{f}$ of $\mathfrak{g}$ such that

$$\mathfrak{g} = \mathfrak{g}' + \mathfrak{f}, \qquad \mathfrak{g}' \cap \mathfrak{f} = \{o\}. \tag{2.1}$$

The tangent space $T_o(F)$ to F at o can be identified with $\mathfrak{f}$; let p be the natural projection of G onto $F = G/G'$, then δp maps $T_e(G)$ onto $T_o(F)$, and since $T_e(G)$ and $\mathfrak{g}$ are identified, δp maps $\mathfrak{f}$ onto $T_o(F)$ isomorphically.

Each element s of G' induces a linear transformation of $T_o(F)$, which we shall denote by L_s. If $\mathfrak{f}$ satisfies

$$ad(s)\cdot\mathfrak{f} \subseteq \mathfrak{f} \qquad s \in G', \tag{2.2}$$

then L_s corresponds to $ad(s)$, when we identify $T_o(F)$ with $\mathfrak{f}$.

Now we shall consruct a $T_o(F)$-valued linear differential form θ on P' satisfying the following conditions:

(θ.1) If $\bar{u} \in T(P')$ and $\theta(\bar{u}) = 0$, then $\delta\pi(\bar{u})$is the zero vector, where π is the projection of P' onto M.

$$(\theta.2) \qquad \theta(\bar{u}s) = L_s^{-1}\theta(\bar{u}) \qquad \bar{u} \in T(P'), \quad s \in G'.$$

$$(\theta.3) \qquad \theta(u\bar{s}) = 0 \qquad u \in P', \quad \bar{s} \in T(G').$$

Let $\bar{u}$ be a tangent vector to P' at u. The projection $\pi: P' \to M$ induces the projection $\delta\pi: T(P') \to T(M)$, and $\delta\pi(\bar{u})$ is a vector tangent to M at $\pi(u)$. As the bundle B is *soudé* to M, the vector $\delta\pi(u)$ can be identified with a vector tangent to F_x at $\sigma(x)$, where $x = \pi(u)$. We shall denote by $\bar{u}^*$ this vector tangent to F_x at $\sigma(x)$. The element $u \in P'$ is considered as a mapping of the standard fibre F onto F_x such that $u(o) = \sigma(x)$, where o is the point of F which defined the isotropic group G'. The map u induces the differential map δu of $T(F)$ onto $T(F_x)$. The inverse image $\delta u^{-1}(\bar{u}^*)$ of $\bar{u}^* \in T(F_x)$ by δu is a vector tangent to F at o, which we denote by $\theta(\bar{u})$. Clearly θ is a linear differential form on P'. If $\theta(\bar{u}) = 0$, then $\bar{u}^*$ is the zero vector; consequently $\delta\pi(\bar{u})$ is also the zero vector, which proves the property (θ.1).

Now we shall verify (θ.2).

We see that $\bar{u}s$ is a tangent vector to P' at us. As $\delta\pi(\bar{u}) = \delta\pi(\bar{u}s)$, we have $\bar{u}^* = (\bar{u}s)^*$.

Then

$$\theta(\bar{u}s) = \delta(us)^{-1}\cdot(\bar{u}s)^* = \delta(us)^{-1}\cdot\bar{u}^* = \delta s^{-1}\cdot\delta u^{-1}(\bar{u}^*) = \delta s^{-1}\theta(\bar{u}) = L_s^{-1}\theta(\bar{u}).$$

Finally we shall prove (θ.3). For any $u \in P'$ and $\bar{s} \in T(G')$, $\delta\pi(u\bar{s})$ is the zero vector. From the definition of θ, it is clear that $\theta(u\bar{s}) = 0$.

Suppose our fibre bundle satisfies only the conditions (s.1)–(s.3). We shall prove that, if there exists a $T_o(F)$-valued linear differential form θ on P', which possesses the properties (θ.1)–(θ.3), then the bundle B satisfies also the condition (s.4).

Let $\bar{x}$ be a tangent vector to M at x and $\bar{u}$ be a tangent vector to P' at u such that

$$\delta\pi(\bar{u}) = \bar{x}.$$

Then $\pi(u) = x$. As $\theta(\bar{u})$ is an element of $T_o(F)$ and u is a map of F onto F_x such that $u(o) = \sigma(x)$, the image $\delta u(\theta(\bar{u}))$ of $\theta(\bar{u})$ by the differential of u is a tangent vector to F_x at $\sigma(x)$. Now we shall show that $\delta u(\theta(\bar{u}))$ depends only on $\bar{x}$ and is independent of the choice of $\bar{u}$ such that $\delta\pi(\bar{u}) = \bar{x}$. If $\bar{u}'$ is a tangent vector to P at the same point u such that $\delta\pi(\bar{u}') = \bar{x}$, from the property ($\theta$.3), $\theta(\bar{u}' - \bar{u}) = 0$; hence

$$\theta(\bar{u}') = \theta(\bar{u}), \qquad \delta u(\theta(\bar{u}')) = \delta u(\theta(\bar{u})).$$

If $\bar{u}' = \bar{u}s$ for some $s \in G'$, then $\bar{u}'$ is tangent to P' at us and

$$\theta(\bar{u}') = L_s^{-1}\theta(\bar{u}).$$

Hence

$$\delta(us)\,\theta(\bar{u}') = \delta u \cdot \delta s \cdot L_s^{-1}\theta(\bar{u}) = \delta u \cdot \theta(\bar{u}).$$

This completes the proof, because, for any $\bar{u}' \in T(P')$ such that

$$\delta\pi(\bar{u}') = \delta\pi(\bar{u}),$$

there is an element $s \in G'$ such that $\bar{u}'s$ is tangent at the same point as u and

$$\delta\pi(\bar{u}'s) = \delta\pi(\bar{u}).$$

If the vector subspace $\mathfrak{f}$ of $\mathfrak{g}$ satisfies (2.1) and (2.2), it can be identified with $T_o(F)$. Therefore θ is considered as an $\mathfrak{f}$-valued linear differential form and the property (θ.2) is replaced by

$$(\theta.2') \qquad \theta(\bar{u}s) = s^{-1}\theta(\bar{u})s \qquad \bar{u} \in T(P'), \quad s \in G'.$$

A homogeneous space $F = G/G'$ is called *weakly reductive* **(8)**, if there is a vector subspace $\mathfrak{f}$ of $\mathfrak{g}$ satisfying (2.1) and (2.2).

THEOREM 1. *A fibre bundle B satisfying the condition* (s.1)–(s.3) *is soudé to M, if and only if there exists a $T_o(F)$-valued linear differential form θ on P' possessing the properties* (θ.1)–(θ.3). *If the homogeneous space $F=G/G'$ is weakly reductive then θ is considered as an $\mathfrak{f}$-valued linear differential form and the property* (θ.2) *is replaced by* (θ.2').

Remarks on weakly reductive homogeneous spaces. In either of the following cases, the homogeneous space F is weakly reductive:

(1) G' is compact,

(2) G' is semi-simple and connected,

(3) G' is discrete.

If F is an affine space (resp. Euclidean space) and G is the affine transformation group (resp. the group of motion) of F, then F is weakly reductive.

If $F = G/G'$ is weakly reductive, then there exists an affine connection on F invariant by G **(8)**. Therefore the linear isotropic group G' is isomorphic to the isotropic group G. If F is a real projective space and G is the projective transformation group of F, then $F = G/G'$ is not weakly reductive.

3. Connections of Cartan. We shall use the same notations as in §2. An infinitesimal connection in P is defined by a $\mathfrak{g}$-valued linear differential form $\tilde{\omega}$ on P with

$$(\tilde{c}.1)\qquad \tilde{\omega}(u\bar{s}) = s^{-1}\bar{s} \qquad u \in P, \quad \bar{s} \in T_s(G),$$

$$(\tilde{c}.2)\qquad \tilde{\omega}(\bar{u}s) = s^{-1}\tilde{\omega}(\bar{u})s \qquad u \in T(P), \quad s \in G.$$

The meaning of $s^{-1}\bar{s}$ and $s^{-1}\tilde{\omega}(\bar{u})s$ is explained in §1.

Let ω be the restriction of the form $\tilde{\omega}$ on P'. Then ω is also a $\mathfrak{g}$-valued linear differential form such that

$$(c.1)\qquad \omega(u\bar{s}) = s^{-1}\bar{s} \qquad u \in P', \quad s \in T_s(G'),$$

$$(c.2)\qquad \omega(\bar{u}s) = s^{-1}\omega(\bar{u})s \qquad u \in T(P'), \quad s \in G'.$$

The form ω does not give a connection in P', because it is not $\mathfrak{g}'$-valued. It is clear that, if ω is a $\mathfrak{g}$-valued linear differential form on P' satisfying the conditions (c.1) and (c.2), then it is the restriction of a unique differential form $\tilde{\omega}$ on P satisfying the conditions (c.1) and (c.2).

An infinitesimal connection in P defined by $\tilde{\omega}$ is called a *connection of Cartan* **(3)**, if the restricted form ω satisfies the following condition:

(c.3) If $\bar{u} \in T(P')$ and $\omega(\bar{u}) = 0$, then $\bar{u}$ is the zero vector. This implies that ω defines an absolute parallelism on P'.

Suppose the homogeneous space $F = G/G'$ is weakly reductive, and let ω' be a $\mathfrak{g}'$-valued linear differential form on P', which defines an infinitesimal connection in P'. The form ω' satisfies the same conditions (c.1) and (c.2) as the form ω; the difference is that the one is $\mathfrak{g}'$-valued and the other is $\mathfrak{g}$-valued. Let θ be the $\mathfrak{f}$-valued linear differential form on P' in Theorem 1. We shall show that the sum $\theta + \omega'$ satisfies the conditions (c.1)–(c.3). Put

$$(3.1)\qquad \omega = \theta + \omega'.$$

Then

$$(3.2)\qquad \omega(u\bar{s}) = \theta(u\bar{s}) + \omega'(u\bar{s}) \qquad u \in P', \quad \bar{s} \in T_s(G').$$

From (θ.3), we obtain

$$(3.3)\qquad \omega(u\bar{s}) = \omega'(u\bar{s}) \qquad u \in P', \quad \bar{s} \in T_s(G').$$

As ω' is a form of connection in P', we have

$$(3.4)\qquad \omega'(u\bar{s}) = s^{-1}\bar{s},$$

which proves that ω satisfies (c.1).

We have

$$(3.5)\qquad \omega(\bar{u}s) = \theta(\bar{u}s) + \omega'(\bar{u}s) \qquad \bar{u} \in T(P'), \quad s \in G'.$$

Since ω' is a form of connection in P', we have

$$(3.6)\qquad \omega'(\bar{u}s) = s^{-1}\omega'(\bar{u})s \qquad \bar{u} \in T(P'), \quad s \in G'.$$

From (θ.2′) and (3.6), it follows that

(3.7) $$\omega(\bar{u}s) = s^{-1}\omega(\bar{u})s \qquad \bar{u} \in T(P'), \quad s \in G'.$$

Suppose

(3.8) $$\omega(\bar{u}) = 0,$$

which implies

(3.9) $$\theta(\bar{u}) = 0, \quad \omega'(\bar{u}) = 0.$$

The first means that $\delta\pi(\bar{u})$ is the zero vector, or that the vector $\bar{u}$ is vertical in the sense of Ambrose **(1)**, and the latter implies that the vector $\bar{u}$ is horizontal **(1)** with respect to the connection in P' defined by ω'. Therefore $\bar{u}$ is the zero vector.

We have proved the following

LEMMA 1. *Suppose $F = G/G'$ is weakly reductive. If ω' is a $\mathfrak{g}'$-valued linear differential form on P' defining a connection in P' and θ is an $\mathfrak{f}$-valued linear differential form on P' satisfying the conditions (θ.1), (θ.1′), (θ.3), then the form $\omega = \theta + \omega'$ defines a connection of Cartan in P; that is, ω is the restriction of a form $\bar{\omega}$ on P defining a connection of Cartan in P.*

Now, suppose that ω is a form on P' satisfying the conditions (c.1)–(c.3). Let θ(resp. ω') be the $\mathfrak{f}$ (resp. $\mathfrak{g}'$) component of ω:

(3.10) $$\omega = \theta + \omega',$$

(3.11) $$\theta(\bar{u}) \in \mathfrak{f}, \quad \omega'(\bar{u}) \in \mathfrak{g}' \qquad \bar{u} \in T(P').$$

We shall prove that θ satisfies the conditions (θ.1), (θ.2′), (θ.3) and that ω' defines a connection in P'.

Suppose

(3.12) $$\theta(\bar{u}) = 0.$$

Then

(3.13) $$\omega(\bar{u}) = \omega'(u) \in \mathfrak{g}'.$$

Take an element $\bar{s} \in T_e(G')$ such that

(3.14) $$s = -\,\omega(\bar{u}).$$

($T_e(G')$ was identified with the Lie algebra $\mathfrak{g}'$ of G'.) Then[2]

(3.15) $$\omega(\bar{u}\bar{s}) = \omega(\bar{u}) + \bar{s} = 0.$$

From (c.3), it follows that $\bar{u}\bar{s}$ is the zero vector; hence

(3.16) $$\delta\pi(\bar{u}) = \delta\pi(\bar{u}\bar{s}) = 0,$$

which proves that θ satisfies (θ.1).

---

[2]The conditions (c.1) and (c.2) are equivalent to the following single condition: $\omega(\bar{u}\bar{s}) = s^{-1}\bar{s} + s^{-1}\omega(\bar{u})s$, because $\omega(\bar{u}\bar{s}) = \omega(u\bar{s}) + \omega(\bar{u}s)$. Putting $s = e$, we obtain (3.15).

Since $\omega(u\bar{s}) = s^{-1}\bar{s}$ is contained in $\mathfrak{g}'$, $\theta(u\bar{s})$ vanishes for any $u \in P'$ and $\bar{s} \in T_e(G')$; hence

(3.17) $$\omega(u\bar{s}) = \omega'(u\bar{s}) \qquad u \in P', \quad \bar{s} \in T_e(G').$$

Therefore θ satisfies (θ.3) and ω' satisfies (c.1). We have

(3.18) $$\omega(\bar{u}s) = s^{-1}(\theta(\bar{u}) + \omega'(\bar{u}))s = s^{-1}\theta(\bar{u})s + s^{-1}\omega'(\bar{u})s \qquad \bar{u} \in T(P'),\ s \in G'.$$

As the homogeneous space F is weakly reductive, $s^{-1}\theta(\bar{u})s$ is contained in $\mathfrak{f}$. Comparing (3.18) with the following equality

(3.19) $$\omega(\bar{u}s) = \theta(\bar{u}s) + \omega'(\bar{u}s),$$

we obtain

(3.20) $$\theta(\bar{u}s) = s^{-1}\theta(\bar{u})s, \qquad \omega'(\bar{u}s) = s^{-1}\omega'(\bar{u})s.$$

Therefore θ satisfies (θ.2′) and ω' satisfies (c.2).

LEMMA 2. *If a $\mathfrak{g}$-valued linear differential form ω on P' satisfies the conditions* (c.1)–(c.3), *then ω is the direct sum of an $\mathfrak{f}$-valued form θ satisfying* (θ.1), (θ.2′), (θ.3) *and a form ω' defining an infinitesimal connection in P'.*

Theorem 1 justifies the following definition: An $\mathfrak{f}$-valued linear differential form θ is called a *form of "soudure,"* if θ satisfies the conditions (θ.1)–(θ.3).

THEOREM 2. *Suppose $F = G/G'$ is weakly reductive. Then, to every pair of a soudure of B and a connection in P', there corresponds a unique connection of Cartan in P. Conversely, to each connection of Cartan in P, there corresponds a unique pair of a soudure of B and a connection in P'. If we denote by θ, ω', ω a form of soudure, a form of connection in P', a form (restricted on P') of Cartan connection in P respectively, then the correspondence is given by $\omega = \theta + \omega'$.*

The Theorem follows immediately from Lemmas 1 and 2.

4. Structure equations. Let ω be a form on P defining a connection of Cartan in P. Then we have

(4.1) $$d\hat{\omega} = -\tfrac{1}{2}[\hat{\omega}, \hat{\omega}] + \hat{\Omega},$$

where $\hat{\Omega}$ is the curvature form (**1**; **3**).

Consider the restricted form ω on P'. Then we have

(4.2) $$d\omega = -\tfrac{1}{2}[\omega, \omega] + \Omega,$$

where Ω is the restriction of $\hat{\Omega}$ on P'.

Assuming the homogeneous space $F = G/G'$ is weakly reductive, we substitute $\omega = \theta + \omega'$ in (4.2) and we obtain

(4.3) $$d\theta + d\omega' = -\tfrac{1}{2}([\theta, \omega'] + [\omega', \theta] + [\omega', \omega'] + [\theta, \theta]) + \Omega.$$

We decompose $[\theta, \theta]$ and Ω into two components as follows:

$$[\theta, \theta] = [\theta, \theta]_{\mathfrak{f}} + [\theta, \theta]_{\mathfrak{g}'}, \quad \Omega = \Omega_{\mathfrak{f}} + \Omega_{\mathfrak{g}'},$$

where, for any $\tilde{u}, \tilde{u}' \in T(P')$ tangent at the same point,

$$[\theta(\tilde{u}), \theta(\tilde{u}')]_{\mathfrak{f}} \in \mathfrak{f}, \qquad [\theta(\tilde{u}), \theta(\tilde{u}')]_{\mathfrak{g}'} \in \mathfrak{g}'$$
$$\Omega_{\mathfrak{f}}(\tilde{u}, \tilde{u}') \in \mathfrak{f}, \qquad \Omega_{\mathfrak{g}'}(\tilde{u}, \tilde{u}') \in \mathfrak{g}'.$$

Then we obtain from (4.3) the following equalities:

$$d\theta = -\tfrac{1}{2}([\theta, \omega'] + [\omega', \theta] + [\theta, \theta]_{\mathfrak{f}}) + \Omega_{\mathfrak{f}}. \tag{4.4}$$

$$d\omega' = -\tfrac{1}{2}([\omega', \omega'] + [\theta, \theta]_{\mathfrak{g}'}) + \Omega_{\mathfrak{g}'}. \tag{4.5}$$

Putting

$$\Theta = \Omega_{\mathfrak{f}} - \tfrac{1}{2}[\theta, \theta]_{\mathfrak{f}}, \tag{4.6}$$

we call Θ the *torsion form* of the connection of Cartan. As the curvature form Ω' of the connection in P' defined by ω' is given by

$$d\omega' = -\tfrac{1}{2}[\omega', \omega'] + \Omega', \tag{4.7}$$

we obtain from (4.5) the following equality.

$$\Omega' = \Omega_{\mathfrak{g}'} - \tfrac{1}{2}[\theta, \theta]_{\mathfrak{g}'}. \tag{4.8}$$

Now we obtain the following

THEOREM 3. *Let*

$$\Theta = \Omega_{\mathfrak{f}} - \tfrac{1}{2}[\theta, \theta]_{\mathfrak{f}}$$

be the torsion form and Ω' the curvature form of the connection in P' defined by ω'. Then we have

$$d\theta = -\tfrac{1}{2}([\theta, \omega'] + [\omega', \theta]) + \Theta,$$
$$\Omega' = \Omega_{\mathfrak{g}'} - \tfrac{1}{2}([\theta, \theta]_{\mathfrak{g}'}).$$

(1) *If the homogeneous space $F = G/G'$ satisfies furthermore the condition*

$$[\mathfrak{f}, \mathfrak{f}] \subseteq \mathfrak{g}',$$

then we have

$$\Theta = \Omega_{\mathfrak{f}}, \qquad \Omega' = \Omega_{\mathfrak{g}'} - \tfrac{1}{2}[\theta, \theta].$$

(2) *If the homogeneous space $F = G/G'$ satisfies the stronger condition*

$$[\mathfrak{f}, \mathfrak{f}] = 0,$$

then we have

$$\Theta = \Omega_{\mathfrak{f}}, \qquad \Omega' = \Omega_{\mathfrak{g}'}.$$

Remarks. A homogeneous space F is called *symmetric*, if it satisfies the assumption of (1) in Theorem 3. On such a space F, there exists (8) an affine

symmetric connection invariant under G. If F is an affine space and G is the affine transformation group, then F satisfies the assumption of (2) in Theorem 3. In this case, a connection in P' is called a *linear connection* (because the structure group G' is the general linear group). If F is an affine space and B is the tangent bundle $T(M)$, then there is a canonical soudure in B. If we take always this canonical soudure, then Theorem 2 says that, to each linear connection in P', there corresponds a unique connection of Cartan in P, which will be called an *affine connection*. Part (2) of Theorem 3 implies that the restriction on P' of the curvature form of an affine connection is the sum of the torsion form and the curvature form of the corresponding linear connection (which is usually called the curvature form of the affine connection).

It will not be useless to point out that the holonomy group of the linear connection corresponding to an affine connection is usually called the homogeneous holonomy group of the affine connection. If the torsion form of an affine connection vanishes, then the form $\Omega_{\mathfrak{f}}$ vanishes also ((2) of Theorem 3). But this does not imply that the form $\hat{\Omega}_{\mathfrak{f}}$, $\mathfrak{f}$-component of the curvature form of the affine connection (of which $\Omega_{\mathfrak{f}}$ is the restriction on P') vanishes. That is why the holonomy group of an affine connection without torsion contains the translation part **(7)**. And we shall see easily that, if the non-homogeneous holonomy group coincides with the homogeneous holonomy group, then our affine connection is flat.

5. Invariant connections of Cartan. Consider a homogeneous space $F = G/G'$. G is considered as a principal fibre bundle over the base manifold F, with structure group G' and with the natural projection **(9)**

$$\pi: G \to F = G/G'.$$

Let P be the fibre bundle with fibre G (on which G' acts on the left)associated to the principal fibre bundle G described above. P is defined as follows. We shall say two elements (s_1, s_2) and (s_3, s_4) of $G \times G$ are equivalent if there is an element s' of G' such that

$$s_1 s' = s_3, \qquad s'^{-1} \cdot s_2 = s_4. \tag{5.1}$$

P is the set of these equivalence classes with the natural structure of fibre bundle; the projection of P onto the base manifold F is induced from the mapping of $G \times G$ onto F:

$$(s_1, s_2) \to \pi(s_1), \tag{5.2}$$

where π is the natural projection of G onto F. The operation of G on $G \times G$ on the right given by

$$(s_1, s_2)\, s = (s_1, s_2 s) \tag{5.3}$$

induces the operation of G on P on the right. In this way, P can be considered as a principal fibre bundle with group G.

The injection of G into $G \times G$ such that $s \to (s, e)$, where e is the unit of G, defines the injection of G into P. The submanifold G of P is stable under the operation of G' on the right; that is, if $u \in P$ belongs to the submanifold G, then us belongs to G for any $s \in G'$.

LEMMA 3. *The princpal fibre bundle P is trivial; P is the direct product of the base space F and the structure group G.*

Proof. Define a mapping j of $G \times G$ onto $F \times G$ as follows:

$$j(s_1, s_2) = (\pi(s_1), s_1 s_2). \tag{5.4}$$

Then j induces a mapping j^0 of P onto $F \times G$, which commutes obviously with the operation of G on the right, proving the Lemma.

As P is trivial, the fibre bundle B with fibre F associated to the principal fibre bundle P is also trivial:

$$B = F \times F. \tag{5.5}$$

LEMMA 4. *The fibre bundle B with fibre F associated to P is soudé* (3) *to the base manifold F.*

Proof. The conditions (s.1) and (s.2) of §1 are apparently satisfied. We take the cross-section σ defined as follows:

$$F \ni x \to (x, x) \in F \times F = B. \tag{5.6}$$

The identification of $T(F)$ with $T_F(B)$ is given by

$$T(F) \ni \bar{x} \to (x, \bar{x}) \in T_F(B). \tag{5.7}$$

If we reduce the structure group G of P to G', we obtain the principal fibre bundle G, from which we started.

The fibre bundle G corresponds to the fibre bundle P' in §2. Therefore *we denote by P' the fibre bundle G.*

A connection of Cartan in P is given by a $\mathfrak{g}$-valued linear differential form ω on $P'(=G)$ satisfying the conditions (c.1)–(c.3). As P' is a group space G, G acts on P' on the left as well as on the right. We shall define a left invariant connection of Cartan; that is, we shall define a $\mathfrak{g}$-valued form ω on P' such that

$$\omega(s\bar{u}) = \omega(\bar{u}) \qquad \bar{u} \in T(P'), \quad s \in G. \tag{5.8}$$

It is clear that such a form ω is unique and must be defined by

$$\omega(u\bar{s}) = \bar{s} \qquad u \in P', \quad \bar{s} \in T_e(G). \tag{5.9}$$

In this case the structure equation of E. Cartan reduces to the equation of Maurer-Cartan:

$$d\omega = -\tfrac{1}{2}[\omega, \omega]. \tag{5.10}$$

THEOREM 4. *There is a unique left invariant connection of Cartan in P. It is given by a $\mathfrak{g}$-valued form ω on $P'(=G)$ defined as follows:*

$$\omega(u\bar{s}) = \bar{s} \qquad u \in P', \quad \bar{s} \in T_e(G).$$

The curvature form of the connection vanishes on P', hence on P, too.

Proof. From (5.10), it follows that the curvature form vanishes on P'. Let $\hat{\Omega}$ be the curvature form. Then we have

$$\hat{\Omega}(\bar{u}s, \bar{u}'s) = s^{-1}\hat{\Omega}(\bar{u}, \bar{u}')s \qquad \bar{u}, \bar{u}' \in T_u(P), \quad s \in G. \tag{5.11}$$

Since $\hat{\Omega}$ vanishes on P', it follows easily from (5.11) that $\hat{\Omega}$ vanishes on P.

Suppose the homogeneous space $F = G/G'$ is weakly reductive. Let

$$\omega = \theta + \omega' \tag{5.12}$$

be the decomposition of the form ω into an $\mathfrak{f}$-valued form θ and into a $\mathfrak{g}'$-valued form ω'. The $\mathfrak{g}'$-valued form ω' defines a connection in the principal fibre bundle $P'(=G)$ with group G'. Let Θ be the torsion form of the connection of Cartan defined by ω and Ω' the curvature form of the connection in P' defined by ω'. From Theorems 3 and 4, it follows that

$$\Theta = -\tfrac{1}{2}[\theta, \theta]_{\mathfrak{f}}, \tag{5.13}$$

$$\Omega' = -\tfrac{1}{2}[\theta, \theta]_{\mathfrak{g}'}. \tag{5.14}$$

THEOREM 5. *Let ω be the $\mathfrak{g}$-valued form on $P'(=G)$ defining the invariant connection of Cartan in P. Suppose the homogeneous space $F = G/G'$ is weakly reductive and let $\omega = \theta + \omega'$ be the decomposition corresponding to a decomposition of the Lie algebra $\mathfrak{g}$ satisfying* (2.2). *Then*

(1) *The torsion form of the connection of Cartan defined by ω is given by*

$$\Theta = -\tfrac{1}{2}[\theta, \theta]_{\mathfrak{f}}.$$

(2) *The curvature form of the connection in P' defined by ω' is given by*

$$\Omega' = -\tfrac{1}{2}[\theta, \theta]_{\mathfrak{g}'}.$$

(3) *The torsion form vanishes, if and only if the homogeneous space F is symmetric; that is,*

$$[\mathfrak{f}, \mathfrak{f}] \subseteq \mathfrak{g}'.$$

(4) *The restricted holonomy group of the connection defined by ω' is an invariant subgroup of the connected component of the unit of G'. And the Lie algebra of the holonomy group is the linear closure of*

$$\{[f_1, f_2]_{\mathfrak{g}'};\ f_1, f_2 \in \mathfrak{f}\}.$$

Proof. We have only to prove (3) and (4). From (θ.1) it follows that, for any f_1, f_2 and $u \in P'$, there are $\bar{u}_1, \bar{u}_2 \in T_u(P')$ such that

$$\theta(\bar{u}_1) = f_1, \qquad \theta(\bar{u}_2) = f_2. \tag{5.15}$$

Therefore, in order that the homogeneous space F be symmetric, it is necessary that the torsion form vanishes. It is evident that, if F is symmetric, the torsion form vanishes.

Now we shall prove (4). Take an arbitrary point u_0 in P' and let P^0 be the set of all points in P' which can be joined to u_0 by horizontal curves **(1)** (with respect to the connection defined by ω'). In other words, we reduce the structure group of P' to the holonomy group of the connection defined by ω', and we obtain the principal fibre bundle P^0 whose structure group is the holonomy group. Then the Lie algebra of the holonomy group is the linear closure of **(1)**.

$$(5.16) \qquad \{\Omega'(\tilde{u}_1, \tilde{u}_2);\ \tilde{u}_1, \tilde{u}_2 \in T_u(P^0),\ u \text{ running through } P^0\},$$

which is equal to

$$(5.17) \qquad \{[\theta(\tilde{u}_1), \theta(\tilde{u}_2)]_{\mathfrak{g}'};\ \ \tilde{u}_1, \tilde{u}_2 \in T_u(P^0)\}.$$

Since, for any $f_1, f_2 \in \mathfrak{f}$ and $u \in P^0$, there are $\tilde{u}_1, \tilde{u}_2 \in T_u(P^0)$ satisfying (5.15), the set (5.17) is equal to the set

$$(5.18) \qquad \{[f_1, f_2]_{\mathfrak{g}'};\ \ f_1, f_2 \in \mathfrak{f}\}.$$

Using the Jacobi's identity, we see easily that the linear closure of the set (5.15) is an ideal of the Lie algebra $\mathfrak{g}'$ of G'. This completes the proof of (4).

Remark. The results in this section are closely related to those of Nomizu on invariant affine connections **(8)**. The relation between them will be discussed in another paper.

References

1. W. Ambrose and I. M. Singer, *A theorem on holonomy*, Trans. Amer. Math. Soc. *75* (1953), 428–443.
2. C. Chevalley, *Theory of Lie groups* (Princeton, 1946).
3. C. Ehresmann, *Les connexions infinitésimales dans un espace fibré différentiable*, Colloque de Topologie (Bruxelles, 1950).
4. S. Kobayashi, *Le groupe des transformations qui laissent invariant le parallélisme*, Colloque de Topologie (Strasbourg, 1954).
5. ———, *Espaces à connexion de Cartan complets*, Proc. Jap. Acad. *30* (1954), 709–710.
6. ———, *Espaces à connexions affines et riemanniennes symétriques*, Nagoya Math. J., *9* (1955) 25–27.
7. S. Kobayashi and K. Nomizu, *Curvature, torsion and holonomy* (unpublished).
8. K. Nomizu, *Invariant affine connections on homogeneous spaces*, Amer. J. Math. *76* (1954), 33–65.
9. N. Steenrod, *The topology of fibre bundles* (Princeton, 1951).

University of Washington, Seattle

A GENERALIZATION OF AN INEQUALITY OF HARDY AND LITTLEWOOD

K. T. SMITH

1. Introduction. A well-known inequality of Hardy-Littlewood reads as follows **(4)**: if $p > 1$ and $f > 0$, then

$$\int_a^b \bar{f}(x)^p dx < A \int_a^b f(x)^p dx,$$

where $\bar{f}(x)$ is defined as the supremum of the numbers

$$\frac{1}{v+u} \int_{x-u}^{x+v} f(t)\, dt;$$

the constant depends on p only. The statement obtained by putting $p = 1$ is false; its substitute reads:

$$\int_a^b \bar{f}(x)\, dx < A \int_a^b f(x)\, dx + B \int_a^b f(x) \log^+ f(x)\, dx + \epsilon;$$

the constants depend on ϵ but not on f. The Hardy-Littlewood inequality has had several important applications: to function theory, harmonic functions, Fourier series, and the strong differentiability of multiple integrals—to mention those with which the author is acquainted. The application to harmonic functions is the following **(4)**:

Let $u(r, \phi)$ be a non-negative harmonic function in the unit circle, and for each ϕ define

$$\bar{u}(\phi) = \sup_{0<r<1} u(r, \phi).$$

Then if $p > 1$,

$$\int_0^{2\pi} \bar{u}(\phi)^p d\phi < A \sup_{0<r<1} \int_0^{2\pi} u(r, \phi)^p d\phi = A \int_0^{2\pi} u(\phi)^p d\phi,$$

where $u(\phi) = u(1, \phi)$ is the boundary function for u and where A is a constant depending on p only.

Since the original appearance of the inequality there have been a number of generalizations. It was formulated for n-dimensional space by Wiener **(14)** and used to prove dominated individual ergodic theorems. The n-dimensional case was used also by Calderón and Zygmund **(3)** to prove dominated pointwise convergence of singular integrals. It was formulated for certain types of

Received February 28, 1955. Work performed under contract with Office of Naval Research N58304.

locally compact topological groups with Haar measure by Calderón (2) and used again to prove ergodic theorems. It was formulated in a weaker version for metric spaces with an outer measure of the type considered below by Rauch (8) and used to prove ergodic theorems and theorems about analytic functions of several complex variables. The latter are described after Theorem 4 below.[1]

The object of this note is to give the inequality a general form valid for certain types of measures on metric spaces and to give applications of the general form of the inequality to harmonic functions, subharmonic functions, and strong differentiability of multiple integrals.

2. The inequality. Most of the arguments are based upon a simple covering theorem which appears implicitly in Banach's proof of the Vitali covering theorem. It can be stated as follows.[2]

LEMMA 1. *Let $\mathfrak{S}$ be a family of spheres in a metric space. If $\mathfrak{S}$ satisfies the conditions* (i) *and* (ii) *below, then it contains a disjoint sequence $\{S(x_n, r_n)\}$ such that*

$$\sum_{S(x,r)\in\mathfrak{S}} S(x, r) \subset \sum_{n=1}^{\infty} S(x_n, 4r_n).$$

The conditions are as follows:

(i) *There is a number R such that for every $S(x, r) \in \mathfrak{S}$, $0 < r < R$.*

(ii) *If $\{S(x_n, r_n)\}$ is any disjoint sequence in $\mathfrak{S}$, then $r_n \to 0$.*

By using the notation $S(x, r)$ for the sphere with center x and radius r we agree tacitly that a sphere is an object determined by a center and a radius. In the applications of Lemma 1 it is the set of points included in the sphere which is important. In order to apply the Lemma to a family of sets each of which is a sphere with respect to several centers and several radii it will be necessary to demonstrate the possibility of choosing for each set one center and one radius in such a way that the hypotheses of the Lemma are satisfied. This arrangement has been picked in order to avoid an unnecessary hypothesis excluding isolated points in the metric space. When the space consists solely of a finite number of isolated points the inequality becomes an inequality on finite sums of some interest in itself. It is to this special case that the greater part of the proof of Hardy and Littlewood is devoted.

We shall consider a metric space B on which there is a regular outer measure subject to the conditions which follow.

[1]Until told by the referee, the author was not aware of the work of Wiener, Calderón, and Rauch. Recently Rauch has supplemented his note (8) with a paper (9) which will be found elsewhere in this journal; in the latter he obtains the full Hardy-Littlewood inequality by a method of Wiener, but with less precise constants than those in the theorems below.

[2]Wiener, Calderón, and Rauch use similar covering theorems. A proof is given in (1).

If E is any set, $|E|$ is its measure and $\delta(E)$ is its diameter.

(a) *Each sphere is measurable and has finite measure.*

(b) *There is a constant K such that $|S(x, 4r)| \leq K|S(x, r)|$ for every closed sphere $S(x, r)$.*

(c) *If $\{S_n\}$ is a sequence of closed spheres such that $|S_n| \to 0$, then*[3] *$\delta(S_n) \to 0$.*

(d) *If $\{S_n\}$ is a sequence of closed spheres such that $\delta(S_n) \to \infty$, then $|S_n| \to \infty$.*

It is known that these conditions are sufficient to ensure that every Borel set in B is measurable.

When f is a non-negative measurable function belonging to some class L^p, $p \geq 1$, on B we make use of the following notations:

(i) $\bar{f}(x)$ is the supremum of the averages of f over all the closed spheres centered at x; that is,

$$\bar{f}(x) = \sup \frac{1}{|S|} \int_S f(y)\, dy,$$

the supremum being taken over all closed spheres S centered at x.

(ii) $f^*(t)$, defined for t real and >0, is the non-increasing equimeasurable rearrangement of f. (that is, $|E_t[f^*(t) > a]| = |E_t[f(x) > a]|$ for all $a > 0$.) It is well known that for any measurable set $E \subset B$,

$$\int_E f(x)\, dx \leq \int_0^{|E|} f^*(t)\, dt,$$

and equality holds if $E = B$.

(iii)
$$\beta_f(t) = \frac{1}{t} \int_0^t f^*(s)\, ds.$$

β_f is a continuous non-increasing function, strictly decreasing except possibly in an interval beginning with 0 where it can be constant.

(iv) β^f is the (upper semi-continuous) inverse function to β_f; if $s > \sup \beta_f(t)$, then[4] $\beta^f(s) = 0$.

LEMMA 2. *If $f \geq 0$ belongs to L^p, $p \geq 1$, then $\bar{f}$ is lower semi-continuous. Consequently $\bar{f}$ is measurable.*

Proof. Suppose that $x_n \to x$. If S is an arbitrary closed sphere with center x and radius r, let S_n be the closed sphere with center x_n and radius $r + d(x_n, x)$. Since $S = \lim S_n$, it follows both that $|S| = \lim |S_n|$ and that

$$\int_S f(y)\, dy = \lim \int_{S_n} f(y)\, dy$$

---

[3]The author's original condition was somewhat less general. The change to this condition and a modification in the proof of Theorem 1 required by the change were suggested by N. Aronszajn.

[4]The properties of f^*, β_f, and β^f are described briefly in Calderón and Zygmund (3).

and hence that

$$\frac{1}{|S|}\int_S f(y)\,dy = \lim \frac{1}{|S_n|}\int_{S_n} f(y)\,dy \leq \liminf \bar{f}(x_n).$$

THEOREM 1. *If $f \geq 0$ belongs to L^p, $p > 1$, then $\bar{f}$ also belongs to L^p and*

$$\int \bar{f}(x)^p dx \leq K\left(\frac{p}{p-1}\right)^p \int f(x)^p dx$$

where K is the constant of hypothesis (b) *on B.*

Proof. For the first part of the argument we suppose only $p \geq 1$. We begin by noting that if E is any measurable set of positive measure over which the average of f is $\geq t > 0$, then by Hölder's inequality

$$t \leq \frac{1}{|E|}\int_E f(x)\,dx \leq \frac{1}{|E|^{1/p}}\left\{\int f(x)^p dx\right\}^{1/p},$$

so that

$$|E| \leq \frac{1}{t^p}\int f(x)^p dx.$$

That is, $|E|$ is bounded by a constant independent of E. If $\{E_n\}$ is a disjoint sequence of sets over which the average of f is $\geq t$, then

$$E = \sum_{n=1}^{\infty} E_n$$

is also a set over which the average of f is $\geq t$. Consequently $\sum |E_n| = |E| < \infty$, so that $|E_n| \to 0$.

Now, if $t > 0$, let B_t denote the set of points x such that $\bar{f}(x) > t$. For each $x \in B_t$, t fixed, let S_x be a closed sphere centered at x over which the average of f is $\geq t$. Furthermore, choose S_x with positive measure and so that it admits a non-zero radius. Let $\mathfrak{S}$ be the family of these spheres. It will be shown that it is possible to choose for each $x \in B_t$ a radius $r(x)$ such that $S_x = S(x, r(x))$ and such that with this choice of centers and radii for the spheres in $\mathfrak{S}$, $\mathfrak{S}$ satisfies the conditions of Lemma 1.

For each $x \in B_t$ let $r_0(x)$ be the infimum and $r_1(x)$ the supremum of the numbers r such that $S_x = S(x, r)$. If $r_0(x) \geq |S_x|$, then take $r(x) = r_0(x)$. If $r_0(x) < |S_x|$, then take

$$r(x) = \min\left[\frac{r_0(x) + r_1(x)}{2}, |S_x|\right].$$

Clearly $r(x) \neq 0$ and $S_x = S(x, r(x))$. From the first paragraph of the proof it follows that the numbers $|S_x|$ for $S_x \in \mathfrak{S}$ are bounded, and then from condition (d) on B it follows that the numbers $\delta(S_x)$ for $S_x \in \mathfrak{S}$ are bounded. Now, either $r(x) = r_0(x) \leq \delta(S_x)$ or $r(x) \leq |S_x|$, so the numbers $r(x)$ for $S_x \in \mathfrak{S}$ are bounded; and condition (i) of Lemma 1 is verified. If

$$\{S_{x_n}\}$$

is a disjoint sequence in $\mathfrak{S}$, then again from the first paragraph it follows that

$$|S_{x_n}| \to 0,$$

and from condition (c) on B it follows that

$$\delta(S_{x_n}) \to 0.$$

Thus $r(x_n) \to 0$, and condition (ii) of Lemma 1 is verified.

Having verified (i) and (ii), we can apply Lemma 1 to extract from $\mathfrak{S}$ a disjoint sequence $\{S(x_n, r_n)\}$ such that

$$B_t \subset \sum_{S(x,r)\in\mathfrak{S}} S(x, r) \subset \sum_{n=1}^{\infty} S(x_n, 4r_n).$$

Then

$$|B_t| < \sum_{n=1}^{\infty} |S(x_n, 4r_n)| < K \sum_{n=1}^{\infty} |S(x_n, r_n)| = K|E|,$$

where K is the constant in hypothesis (b), and $E = \sum S(x_n, r_n)$. As before, the average of f over E is $>t$. (Therefore the first paragraph provides a bound for $|E|$, but this bound is not sharp enough.) We have however,

$$t < \frac{1}{|E|}\int_E f(x)\,dx < \frac{1}{|E|}\int_0^{|E|} f^*(t)\,dt = \beta_f(|E|),$$

and by inverting β_f, $|E| < \beta^f(t)$. Finally, therefore, $|B_t| < K\beta^f(t)$.

Now let $p > 1$. In the following chain of inequalities we use the fact that

$$\lim_{s\to\infty} s\beta_f(s)^p = \lim_{s\to 0} s\beta_f(s)^p = 0,$$

and we use the substitution $t = \beta_f(s)$. We have[5]

$$\begin{aligned}\int \bar{f}(x)^p dx &= \int_0^{\infty} pt^{p-1}|B_t|\,dt < K\int_0^{\infty} pt^{p-1}\beta^f(t)\,dt \\ &= -K\int_0^{\infty} s\,d\beta_f(s)^p = K\int_0^{\infty}\beta_f(s)^p ds = K\int_0^{\infty}\frac{1}{s^p}\left\{\int_0^s f^*(t)\,dt\right\}^p ds \\ &< K\left(\frac{p}{p-1}\right)^p\int_0^{\infty} f^*(t)^p dt = K\left(\frac{p}{p-1}\right)^p\int_B f(x)^p dx.\end{aligned}$$

THEOREM 2. *If f is non-negative and measurable, and if $f(x)\log^+ f(x)$ is integrable, then for any measurable set E,*

$$\int_E \bar{f}(x)\,dx < 2K\int f(x)\log^+ f(x)\,dx + \left(\frac{4K}{e} + 2\right)|E|.$$

[5] This concluding calculation can be found in Calderón and Zygmund (3). The last inequality in the chain is a well-known inequality of Hardy.

Proof. It is well known that if $f \log^+ f$ is integrable, then f itself is integrable over every set of finite measure. This is all that is necessary to the formation of the function $\bar{f}$. It does not guarantee the existence of f^*, however, so we write $f = g + h$, where $g(x) = f(x)$ if $f(x) < 2$ and $g(x) = 0$ otherwise.

It is clear that $\int_E \bar{g}(x)\, dx < 2\,|E|$, so if it can be proved that

$$\int_E \bar{h}(x)\, dx < 2K \int h(x) \log^+ h(x)\, dx + \frac{4K}{e}|E|, \tag{2.1}$$

then we will have

$$\int_E \bar{f}(x)\, dx < \int_E \bar{g}(x)\, dx + \int_E \bar{h}(x)\, dx < 2K \int f(x) \log^+ f(x)\, dx + \left(\frac{4K}{e} + 2\right)|E|.$$

Now, h is in fact an integrable function, so the proof will be complete if we prove (2.1) for integrable functions.

Let us call the integrable function f, rather than h, so that the notations used in Theorem 1 will be appropriate. Let $B_t' = E \cap B_t$. Then $|B_t'| < |E|$ and, as was proved in Theorem 1, $|B_t'| < |B_t| < K\beta^f(t)$. Hence

$$\int_E \bar{f}(x)\, dx = \int_0^\infty |B_t'|\, dt < \int_0^{t_0} |E|\, dt + K \int_{t_0}^\infty \beta^f(t)\, dt$$

for any t_0, while also, for $t_0 = \beta_f(|E|)$,

$$\int_{t_0}^\infty \beta^f(t)\, dt = -\int_0^{|E|} s\, d\beta_f(s) = -\,|E|\beta_f(|E|) + \int_0^{|E|} \beta_f(s)\, ds.$$

Furthermore,

$$\int_0^{|E|} \beta_f(s)\, ds = \int_0^{|E|} \frac{ds}{s} \int_0^s f^*(s')\, ds'$$

$$= \int_0^{|E|} f^*(s') \log \frac{|E|}{s'}\, ds' < 2 \int_0^{|E|} f^*(s') \log^+ f^*(s')\, ds' + \frac{2}{e} \int_0^{|E|} \left(\frac{|E|}{s'}\right)^{\frac{1}{2}} ds'$$

$$< 2 \int_E f(x) \log^+ f(x)\, dx + \frac{4}{e}\,|E|.$$

Therefore

$$\int_E \bar{f}(x)\, dx < (1 - K)\,|E|\,\beta_f(|E|) + 2K \int_E f(x) \log^+ f(x)\, dx + \frac{4K}{e}\,|E|.$$

Since necessarily $K \geq 1$, (2.1) follows.

Once the estimate for $|B_t|$ is obtained, the evaluation is almost identical with that given in Calderón and Zygmund **(3)** in the proof of Theorem 2. One of the inequalities used in the chain is that of W. H. Young, namely $ab < a \log a + e^{b-1}$.

THEOREM 3. *If f is non-negative and integrable, and $0 < \epsilon < 1$, then for every measurable set E,*

$$\int_E \bar{f}(x)^{1-\epsilon} dx < \left(\frac{1}{\epsilon}\right)|E|^\epsilon K^{1-\epsilon} \left\{\int f(x)\, dx\right\}^{1-\epsilon}.$$

Proof. The proof is the same as the last part of the proof of Theorem 3 of Calderón and Zygmund **(3)**. Use must be made, of course, of our previous estimate for $|B_t|$.

3. Applications—Harmonic functions. We propose to apply the inequality to the case where B is a smooth surface bounding a bounded domain D in Euclidean n-dimensional space R^n, $n \geqslant 3$. The explicit smoothness assumptions are as follows.[6]

(a) *B is a C^1 surface; that is, each point of B has an n-dimensional neighborhood V which can be mapped in* 1–1 *fashion on an n-dimensional cube by a transformation T such that T and T^{-1} are C^1 transformations and such that $T(B \cap V)$ is the intersection of the cube with one of the coordinate hyperplanes.*

(b) *B is of bounded curvature in the large sense, that is, if $\alpha(x, y)$ denotes the angle between the exterior normals at x and y, then*

$$\frac{1}{\rho_0} = \sup_{x \neq y} \frac{\sin \frac{1}{2}\alpha(x, y)}{\frac{1}{2}|x - y|} < \infty.$$

The metric in B is its metric as a subset of Euclidean space. The measure on B is the area measure, definable in the classical manner because of the smoothness conditions.

We shall use capital letters P, Q, etc. to designate points in the interior of D, and small letters x, y, etc., to designate points on the boundary B. Each point P at distance less than ρ_0 from B lies on a unique line segment of length less than ρ_0 and normal to B. We write x_P for the point at which this segment meets B. Proceeding from the opposite direction, we write $P(\rho, x)$ for the point at distance ρ from x measured along the interior normal through x. Finally, we write $\mathfrak{F}_p$ for the class of functions $f(P)$ harmonic in D and such that

$$\sup_{0<\rho<\rho_0} \int_B |f[P(\rho, x)]|^p \, dx < \infty.$$

(Note that $f[P(\rho, x)]$ is the restriction of $f(P)$ to the surface parallel to B at distance ρ.)

It is known that if $f \in \mathfrak{F}_p$ then $f(P)$ has a limit $f(x)$ as $P \to x$ non-tangentially **(13)** for almost every point $x \in B$. The so defined function $f(x)$, which belongs to L^p on B, is called the boundary function of f. When $p > 1$, the functions $f[P(\rho, x)]$ converge in mean of order p to the boundary function as $\rho \to 0$, and f is the Poisson integral of the boundary function.[7] We shall make

[6]A forthcoming note by Aronszajn will contain proofs of all the needed properties of such surfaces. The object of his note is to exhibit the best possible constants in all cases. Here we do not need the best constants, but only the qualitative sense of the properties, and for the most part this is classical information.

[7]A proof of the mean convergence can be found in **(1)**. A related result concerning the constant surfaces of the Green's function (for fixed pole) rather than the parallel surfaces is proved by Privaloff and Kouznetzoff **(7)**.

use of an inequality between the values of f in D and the mean values of the boundary function over spheres in B.

Mean value inequality. *If the harmonic function $f(P)$ is the Poisson integral of its boundary function $f(x)$, and if $\bar{f}(x)$ denotes the supremum of the averages of $|f(y)|$ over the closed spheres in B centered at x, then for every P within distance ρ_0 of B we have $|f(P)| < A\,\bar{f}(x_P)$. The constant A depends only on ρ_0 and on the dimension*[8] *n.*

THEOREM 4. *For each f in $\mathfrak{F}_p$ let*

$$\bar{f}(x) = \sup_{0<\rho<\rho_0} |f[P(\rho, x)]|.$$

For $p > 1$, the following assertion holds: if f belongs to $\mathfrak{F}_p$, then $\bar{f}$ belongs to L^p on B, and

$$\int_B \bar{f}(x)^p dx < K\left(\frac{p}{p-1}\right)^p A^p \int_B |f(x)|^p dx = \lim_{\rho\to 0} K\left(\frac{p}{p-1}\right)^p A^p \int_B |f[P(\rho, x)]|^p dx.$$

Proof. The metric space B and its measure are obviously of the type considered in the second section, so Theorem 4 follows directly from Theorem 1 and the mean-value inequality.

In the case of the circle in the plane this is the theorem of Hardy and Littlewood quoted in the introduction. The related theorem of Rauch (**8**; **9**) on analytic functions is as follows:

If D is the sphere, if f is complex valued, and if n is even and the variables can be paired so that f is an analytic function of $n/2$ complex variables, then the assertion of Theorem 4 holds for any exponent $p > 0$.

Rauch's theorem is obtained from the special case of exponent 2 in Theorem 6 below by putting $s(P) = |f(P)|^{\frac{1}{2}p}$.

THEOREM 5. *If f belongs to $\mathfrak{F}_1$, then for each ϵ, $0 < \epsilon < 1$,*

$$\int_B \bar{f}(x)^{1-\epsilon} dx < \infty.$$

Proof. The function $f \epsilon \mathfrak{F}_1$ has a boundary measure ν in terms of which it can be represented as a Poisson-Stieltjes integral. The mean value inequality is valid here in a suitably modified form; namely, $|f(P)|$ is less than or equal to a constant times the upper bound of the quotients $\nu(C)/|C|$ taken over the closed spheres in B centered at x_P. We do not give more of the proof for it is essentially the same as the proof of Theorem 7 below on subharmonic functions.

Remark. The proof of Theorem 5 is not based on Theorem 3, for f does not necessarily have a boundary function of which it is the Poisson integral. It is plain that if f does have such a boundary function, then certain conclusions can be drawn from Theorems 2 and 3. It does not seem necessary to state the conclusions.

[8]This is a special case of an inequality which is proved in Aronszajn and Smith (**1**). This special case was obtained for the circle in the plane by Hardy and Littlewood (**4**).

Subharmonic functions.

THEOREM 6. *Let $s(P)$ be a non-negative subharmonic function in D, and let*

$$\bar{s}(x) = \sup_{0<\rho<\rho_0} s[P(\rho, x)].$$

For $p > 1$, we have[9]

$$\int_B \bar{s}(x)^p dx < K\left(\frac{p}{p-1}\right)^p A^p \sup_{0<\rho<\rho_0} \int_B s[P(\rho, x)]^p dx.$$

Proof. We suppose that the right side is finite. For each ρ, $0 < \rho < \rho_0$, we write B_ρ for the set of points in D at distance ρ from B, and D_ρ for the sub-domain of D bounded by B_ρ. We write f_ρ for the harmonic function in D_ρ with the same boundary values[10] as s. It is well known that the harmonic function f_ρ converge increasingly as $\rho \to 0$ to a harmonic function $f \in \mathfrak{F}_p$ which dominates s **(10)**. In addition the functions $s[P(\rho, x)]$ converge weakly in L^p on B as $\rho \to 0$ to the boundary function $f(x)$ for f. Theorem 6 follows from Theorem 4 and the lower semi-continuity of the norm in L^p with respect to weak convergence.

By using the results of F. Riesz on the representation of subharmonic functions by potentials we can prove similar theorems for subharmonic functions which are not necessarily non-negative. For the sake of simplicity we confine the discussion to the sphere, though the results are equally valid for the more general domains of the last paragraph, as the proofs will show.

The theorem of F. Riesz states that if $s(P)$ is a subharmonic function in the domain D, then a necessary and sufficient condition that $s(P)$ be the sum of a harmonic function and the Green's potential of a negative Borel measure on D is that $s(P)$ be bounded above in D by a harmonic function; the harmonic function which figures in the representation is the smallest harmonic function which bounds $s(P)$ above **(10)**. This function is called the smallest harmonic majorant of $s(P)$. If the positive part of $s(P)$, which we call $s^+(P)$, satisfies the condition

$$\sup_{0<\rho<1} \int_B s^+(\rho\, x)\, dx < \infty,$$

then the smallest harmonic majorant $h(P)$ exists and satisfies

$$\sup_{0<\rho<1} \int_B |h(\rho\, x)|\, dx < \infty.$$

Therefore, as $s(P) = -\int_D G(P, Q)\, d\mu(Q) + h(P)$, where μ is a positive Borel measure on D and $G(P, Q)$ is the Green's function of D; and as $h(P)$ satisfies the hypotheses of Theorem 5, the analogue of Theorem 5 for subharmonic

[9]For the circle in the plane this is a result of Hardy and Littlewood **(4)**.

[10]The surfaces B_ρ are also C^1 and of bounded curvature. The curvature constant ρ_0' for $B_{\rho'}$ is $\rho_0 - \rho'$. f_ρ is defined by the Poisson integral over B_ρ.

functions will result from an analysis of the first term, the Green's potential, alone. Before stating the theorem we observe that the upper bound of $|s(P)|$ along the various radii will be identically infinite whenever the Green's potential is infinite at the origin. Therefore a small sphere with center at 0 must be removed from D before taking the upper bounds.

THEOREM 7. *Let ρ_0 be a fixed number between* 0 *and* 1, *and put*

$$\bar{s}(x) = \sup_{\rho_0<\rho<1} |s(\rho x)|$$

for each $x \in B$. If $s(P)$ is subharmonic in D and if

$$\sup_{0<\rho<1} \int_B s^+(\rho x)\, dx < \infty,$$

then for each ϵ, $0 < \epsilon < 1$,

$$\int_B \bar{s}(x)^{1-\epsilon} dx < \infty.$$

Proof. As we have mentioned, it results from Theorem 5 and the theorem of F. Riesz that we need only prove the theorem for functions of the type $s(P) = -\int_D G(P, Q)\, d\mu(Q) = -u(P)$, where μ is a positive Borel measure on D and $G(P, Q)$ is the Green's function for D. The explicit expression for the Green's function is well known.

$$G(P, Q) = \frac{1}{\omega_n(n-2)}\left[\frac{1}{|P-Q|^{n-2}} - \frac{r^{n-2}}{|Q|^{n-2}}\,\frac{1}{|P-Q'|^{n-2}}\right], \tag{3.1}$$

where

$$Q' = \frac{r^2}{|Q|^2} Q,$$

and ω_n is the area of the surface of the unit sphere.

It is known that the Green's potential $u(P) = \int_D G(P, Q)\, d\mu(Q)$ either is identically $+\infty$ or is finite except at a set of points of outer capacity 0. For the latter to be the case it is necessary and sufficient that [11] $\int_D (r - |Q|)\, d\mu(Q) < \infty$. For each $x \in B$ and each real ξ, $0 < \xi < 2r$, let $C(x, \xi)$ be the sphere in B with center x and radius ξ; and let $S(x, \xi)$ be the conical sector in D generated by joining each point of $C(x, \xi)$ to the origin. Let

$$I(x, \xi) = \int_{S(x,\xi)} (r - |Q|)\, d\mu(Q),$$

and let

$$m(x) = \sup_{\xi} \frac{I(x, \xi)}{|C(x, \xi)|}.$$

[11] This is clear for the sphere. In the case of more general domains the integrand $r - |Q|$ is replaced by $|x_Q - Q|$, the distance from Q to the boundary. In this form the fact was observed by Privaloff and Kousnetzoff (7).

For the present we assume the following Lemma.

LEMMA 3. *There is a constant A such that $\bar{u}(x) < Am(x)$, where $\bar{u}$ is defined like $\bar{s}$.*

The covering theorem is used as in the proof of the general Hardy-Littlewood inequality. Let B_t, $t > 0$, denote the set of points x such that $\bar{u}(x) > t$. If $x \in B_t$, then there is a ξ_x such that

$$\frac{I(x, \xi_x)}{|C(x, \xi_x)|} > \frac{t}{A}.$$

Choosing such a ξ_x for each $x \in B_t$ we have $B_t \subset \sum C(x, \xi_x)$, so by the covering theorem there is a disjoint sequence

$$C(x_n, \xi_n) \ (\xi_n = \xi_{x_n})$$

such that

$$B_t \subset \sum_{n=1}^{\infty} C(x_n, 4\xi_n).$$

If K is chosen so that for all C, $|C(x, 4\xi) < K|C(x, \xi)|$, then [12]

$$|B_t| < \sum_{n=1}^{\infty} |C(x_n, 4\xi_n)| < K \sum_{n=1}^{\infty} |C(x_n, \xi_n)| < \frac{KA}{t} \sum_{n=1}^{\infty} I(x_n, \xi_n)$$
$$= \frac{KA}{t} \int_E (r - |Q|)\, d\mu(Q),$$

where E is the sum of the disjoint sets $S(x_n, \xi_n)$. Hence $|B_t| < k'/t$ for $k' = KA \int_D (r - |Q|)\, d\mu(Q)$. Now

$$\int_B \bar{u}(x)^{1-\epsilon} dx = \int_0^{\infty} (1 - \epsilon)\, t^{-\epsilon} |B_t|\, dt < k'' \int_0^1 t^{-\epsilon} dt + k'(1 - \epsilon) \int_1^{\infty} \frac{dt}{t^{1+\epsilon}}$$

where k'' is larger than $(1 - \epsilon)$ times the area of the surface of the sphere.

Proof of the Lemma. We shall not give the entire proof. The calculations, which are routine, are achieved by majorizing the Green's function ((3.2) below) and considering separately the integrals over three different parts of the sphere. The majoration for the Green's function is obtained by inspection of the explicit formula (3.1).[13]

(3.2) *There is a constant k such that*

$$0 < G(P, Q) < k(r - |P|)(r - |Q|)/ |P - Q|^n;$$

also

$$G(P, Q) < \frac{1}{\omega_n(n - 2)} \frac{1}{|P - Q|^{n-2}}.$$

[12] $|E|$ is used for subsets of D to refer to Lebesgue measure and for subsets of B to refer to the area measure on B.

[13] Essentially the same majoration and division of the sphere are used by Littlewood (6) to prove that in the case of the circle in the plane a Green's potential has radial limit 0 at almost every boundary point. The majoration is valid for any domain bounded by a C^1-surface of bounded curvature (11) and for even more general domains (5).

The division of the sphere is as follows. Let P be fixed with $|P| > \rho_0 r$, let $x = rP/|P|$, and let $\xi_0 = r - |P|$. One part of the sphere is the exterior of the conical sector $S(x, \xi_0)$; another is that part of $S(x, \xi_0)$ whose points Q satisfy $r - |Q| > \frac{1}{2}(r - |P|)$; the third is that part of $S(x, \xi_0)$ whose points Q satisfy $r - |Q| < \frac{1}{2}(r - |P|)$. Finally, it is necessary to use an evaluation of $|P - Q|$ in terms of the variable

$$\xi = \left| \frac{r}{|Q|} Q - \frac{r}{|P|} P \right| .$$

(3.3) *There is a constant k such that if $|P| > \rho_0 r$, then $|P - Q| > k\xi$.*

The Lemma results from simple calculation with these estimates and the remark that the quotient $|C(x, \xi)|/\xi^{n-1}$ is bounded above and from 0.

Theorem 7 can be improved if it is known that the measure is the indefinite integral of a density subject to certain conditions.

THEOREM 8. *If $u(P) = \int_D G(P, Q) f(Q)\, dQ$ where $f(Q)$ is such that*

$$\int_D (r - |Q|)^p f(Q)^p dQ < \infty, \qquad p > 1,$$

then $\bar{u}(x)$ belongs to L^p on B, and there is a constant M such that

$$\int_B \bar{u}(x)^p dx < M^p \int_D (r - |Q|)^p f(Q)^p dQ.$$

Proof. The proof is similar to the proof of the last theorem, but it is possible to make use of the non-increasing rearrangements as in the proof of Theorem 1 in order to obtain better evaluations. With the notations of the last theorem we have, as we had there,

$$|B_t| < K \sum_{n=1}^{\infty} |C(x_n, \xi_n)| = K|C|, \quad C = \sum_{n=1}^{\infty} C(x_n, \xi_n).$$

Because of the disjointness it happens in Theorem 1 that

$$t < \frac{A}{|C|} \int_E (r - |Q|) f(Q)\, dQ$$

(where again $E = \sum S(x_n, \xi_n)$). From the fact that $|E| = r/n|C|$, it follows that

$$t < \frac{A}{|C|} \int_0^{|E|} g^*(s)\, ds = \frac{rA}{n|E|} \int_0^{|E|} g^*(s)\, ds = \frac{rA}{n} \beta_g(|E|)$$

for $g(Q) = (r - |Q|) f(Q)$. Hence

$$|B_t| < K|C| = \frac{Kn}{r} |E| < \frac{Kn}{r} \beta^{\prime}\left(\frac{n}{Ar} t\right).$$

The proof is finished in the same manner as the proof of Theorem 1.

Strong differentiability of double integrals. The general Hardy-Littlewood inequality yields a generalization of the theorem of Jessen, Marcinkiewicz, and Zygmund on the strong differentiability of multiple integrals **(12)**. However, we need the inequality in a slightly stronger form.

Theorems 1, 2, *and* 3 *remain true and their proofs remain correct when* $\bar{f}(x)$ *is redefined to be the supremum of the averages of* $f(y)$ *over all spheres containing* x.

THEOREM 9. *Let* B_1 *and* B_2 *be metric spaces with measures of the kind considered in* §2, *and let* $f(x, y)$ *be a measurable function on* $B_1 \times B_2$. *If*

$$\int_{B_1}\int_{B_2} |f| \log^+ |f| \, dxdy < \infty,$$

then the indefinite integral of f *is almost everywhere derivable in the strong sense. That is, for almost every choice of* (x, y),

$$\lim_n \frac{1}{|S_{1,n}|} \frac{1}{|S_{2,n}|} \int_{S_{1,n}} \int_{S_{2,n}} f(s, t) \, dsdt$$

exists for all sequences $\{S_{1,n}\}$ *and* $\{S_{2,n}\}$ *of closed spheres such that* $x \in S_{1,n}$, $y \in S_{2,n}$, $\delta(S_{1,n}) \to 0$, *and* $\delta(S_{2,n}) \to 0$.

Proof (cf. **12**, pp. 147–149). Several earlier theorems are required (notably, the Vitali covering theorem, the strong density theorem, and the theorem on the strong differentiability of the indefinite integral of a bounded function); these theorems are true in the present case, and the proofs given by Saks are valid after simple modifications.

Remark. It was noticed by Hardy and Littlewood and by Calderón and Zygmund that the Hardy-Littlewood inequality leads to certain results on integral operators. The results are of such a kind as to establish dominated convergence of sequences of transforms. Thus, for example, Hardy and Littlewood show dominated convergence of the Fejer polynomials formed from the Fourier series of a function f; and Calderón and Zygmund show dominated convergence of singular integrals. Our general case of the inequality leads to similar results, which can be used, for example, to give another proof of Theorem 4. However, since we do not have applications which would lead to new results, we shall omit the statement of this theorem on integral operators. In any case it is a re-phrasing in the abstract terms of the theorems of the authors cited.

REFERENCES

1. N. Aronszajn and K. T. Smith, *Functional spaces and functional completion.* To appear shortly in Ann. Inst. Fourier, Grenoble.
2. A. P. Calderón, *A general ergodic theorem.* Ann. Math., *8* (1953), 182–191.
3. A. P. Calderón and A. Zygmund, *On the existence of certain singular integrals*, Acta Math., *88* (1952), 85–139.

4. G. H. Hardy and J. E. Littlewood, *A maximal theorem with function-theoretic applications*, Acta Math., *54* (1930), 81–116.
5. M. Keldych and M. Lavrentieff, *Sur une évaluation pour la fonction de Green*, C.R. Ac. Sci. U.S.S.R., *24* (1939), 22–24.
6. J. E. Littlewood, *On functions subharmonic in a circle*, Lond. Math. Soc., *2* (1927), 192–196.
7. I. I. Privaloff and P. Kouznetzoff, *Sur les problèmes limites et les classes différentes de fonctions harmoniques et subharmoniques définies dans un domaine arbitraire*, Rec. Math. Moscou, *6* (1939), 345–376.
8. H. E. Rauch, *Généralisation d'une proposition de Hardy et de Littlewood et de théorèmes ergodiques qui s'y rattachent*, C.R. Ac. Sci. Paris, *22* (1948), 887–889.
9. ———, *Harmonic and analytic functions of several variables and the maximal theorem of Hardy and Littlewood*, Can. J. Math., *8* (1956), 171–183.
10. F. Riesz, *Sur les fonctions subharmoniques et leur rapport à la théorie du potentiel*, Acta Math., *54* (1930), 321–360.
11. A. Rosenblatt, *Sur la fonction de Green d'un domaine borné de l'espace à trois dimensions*, C.R. Ac. Sci. Paris, *201* (1935), 22–24.
12. S. Saks, *Theory of the Integral* (New York, 1937).
13. C. de la Vallée Poussin, *Propriétés des fonctions harmoniques dans un domaine ouvert limité par des surfaces à courbure bornée*, Ann. Scuola Norm. Sup. Pisa, (2) (1933), 167–197.
14. N. Wiener, *The ergodic theorem*, Duke Math. J., *5* (1939), 1–18.

University of Kansas

HARMONIC AND ANALYTIC FUNCTIONS OF SEVERAL VARIABLES AND THE MAXIMAL THEOREM OF HARDY AND LITTLEWOOD

H. E. RAUCH

1. Introduction and principal theorems. The present paper, an edited excerpt from my dissertation,[1] arose from the suggestion of S. Bochner that I try to extend the maximal theorem of Hardy and Littlewood **(2)** to functions analytic in the solid unit hypersphere

$$S_{2n}\colon\ r^2 \equiv |z_1|^2 + \ldots + |z_n|^2 < 1.$$

If one writes the analytic function of n complex variables, $f(z_1, \ldots, z_n)$, as $f(r, P)$ where

$$P \in S_{2n-1}\colon\ |z_1|^2 + \ldots + |z_n|^2 = 1,$$

then the theorem in question and its generalization are contained in

THEOREM 1. *If, for some $\lambda > 0$, f satisfies*

$$\int_{S_{2n-1}} |f(r, P)|^\lambda dV_P < C^\lambda,\quad r < 1, \tag{1}$$

where dV_P is the volume element on S_{2n-1} at P and C^λ is a constant, then for the same λ

$$\int_{S_{2n-1}} (\sup_{0<r<1} |f(r, P)|)^\lambda dV_P < \alpha_n C^\lambda, \tag{2}$$

α_n being independent of f.

From Theorem 1 one can deduce a generalization of a classical theorem due to the brothers Riesz **(7**, Chap. VII**)**:

THEOREM 2. *Under the same general hypotheses in and preceding Theorem* 1, *and assuming* (1), *there exists a function $f(P)$ of class L^λ on S_{2n-1} such that*

$$\lim_{r\to 1} \int_{S_{2n-1}} |f(r, P) - f(P)|^\lambda dV_P = 0. \tag{3}$$

Received May 6, 1955.

[1]Princeton, 1947. Abstracts of the results appeared as **(3)** and **(4)**. The decision to publish in full, after so long a delay, is motivated by repeated requests of other workers in the field and by overlapping with published material obtained independently by Zygmund, Calderón, and others. A particular impetus is the preceding paper by K. T. Smith **(5)**, in which a substantial part of the underlying methods and results of my paper are obtained independently from a point of view not too different from mine.

As Zygmund remarked **(9)**, Theorem 2 follows immediately from Theorem 1 and the theorem of Calderón and Zygmund to the effect that, under the same hypotheses, $f(r, P)$ has a point-wise limit, $f(P)$, almost everywhere. For the latter, convergence would be majorized according to (2) and would, therefore, imply mean convergence. However, in the less delicate sense $\lambda > 1$ (3) follows directly from (2) without the intervention of the theorem on point-wise convergence, as will be seen in §5.

Now the proof of Theorem 1, to be found in §5, can be reduced by means of a sequence of theorems on analytic, harmonic, and subharmonic functions (§§4 and 5) exactly as in **(2)** to the proof of a theorem of purely real-variable nature:

THEOREM 3. *Let $f(P)$ belong to L^p, $p > 1$, on*

$$S_{n-1}:\quad x_1^2 + \ldots + x_n^2 = 1,$$

and let $\sigma_r(P)$ be the spherical cap of radius r (measured on S_{n-1}) about P on S_{n-1} and $V(r)$ its volume as measured on S_{n-1}. Define $f^*(P)$ by

$$f^*(P) = \sup_{0<r<\pi} \frac{1}{V(r)} \int_{\sigma_r(P)} |f(P')|\, dV_{P'}. \tag{4}$$

Then $f^*(P)$ satisfies

$$\int_{S_{n-1}} \{f^*(P)\}^p dV_P < C_{n,p} \int_{S_{n-1}} |f(P)|^p dV_P, \tag{5}$$

where $C_{n,p}$ depends only[2] *on n and p. If $p = 1$ this is no longer true; however, if $|f(P)|\log^+|f(P)|$ is integrable, then*

$$\int_{S_{n-1}} f^*(P)\, dV_P < B_n \int_{S_{n-1}} |f(P)|\log^+|f(P)|\, dV_P + C_n. \tag{6}$$

The proof of Theorem 3 is the essence of the matter, and the method of analysis was supplied by Wiener in a profound paper **(6)**. There he shows, by a reasoning which is closely related to F. Riesz's proof of the case $n = 1$ of Theorem 3, but simpler and more powerful, that both Birkhoff's ergodic theorem and the Hardy-Littlewood theorem for $n = 1$ have a common source and that both can be extended by the same method, the former to a theorem on averages over an n-parameter abelian group, the latter to a theorem on averages over Euclidean n-space.

Now, Wiener in a lucid fashion reduces everything to a simple measure-theoretic lemma, which he calls "of Vitali type" although it is much more elementary. In studying his paper I noticed that this lemma, although formulated for sets in ordinary n-space, in fact applied to a more general situation from which, in particular, Theorem 3 would follow by Wiener's arguments.

A diagnosis of the elements needed explicitly or implicitly in extending

[2]Wiener's method does not deliver the best constants. Smith's paper **(5)** does.

Wiener's argument to, say, the surface of the hypersphere leads one to describe a metric space with a metric M and an outer measure m as having Euclidean character or *Property* A if, without regard to logical niceties, it is such that (i) spheres of equal radius in M have equal measure in m and vice versa (this very restrictive condition of *homogeneity* may be replaced by a much weaker one of a sort of uniformity in important cases); (ii) countable sets are null-sets; and, most important, (iii) the measure of the set γ covered by a sphere σ and all spheres overlapping σ and having smaller or equal radius satisfies $m(\gamma) < Cm(\sigma)$ where C depends only on M and m.
Then one has

THEOREM A. *In a space possessing Property* A *let a set S of outer measure $m(S)$ be such that every $P \epsilon S$ is the center of one member, $\sigma(P)$, of a certain family of spheres. Then given $\epsilon > 0$ there is a finite number of mutually disjoint members, σ_i, of the family such that*

$$\Sigma_i\, m(\sigma_i) > C^{-1}\, m(S) - \epsilon \tag{7}$$

where C is the constant of Property A.

For $(n-1)$-space and S_{n-1} (as will be seen) $C = 3^{n-1}$. The proof of Theorem A will occupy §2, and the deduction of Theorem 3, §3.

The generality of Theorem A permits the immediate extension of Wiener's generalization of Birkhoff's ergodic theorem to those groups of measure-preserving transformations of a set which admit an invariant metric possessing Property A and which may well be non-commutative. This application is in my dissertation;[3] but I do not reproduce it here since there is already a surfeit of related ergodic theorems on the market.

Besides the hypersphere there are other generalizations of the unit circle, notably the *polycylinder*: $|z_1| = r_1 < 1, \ldots, |z_n| = r_n < 1$ whose boundary is the multitorus,

$$T_n\colon |z_1| = 1, \ldots, |z_n| = 1.$$

The analogue of Theorem 1 for this domain was derived independently and announced almost simultaneously by Zygmund **(8)** in stronger form and me **(3)**. I prove it again here not merely because the proof is different but because the technique of proof will serve to demonstrate a more interesting generalization **(4)**:

THEOREM 4. *Let $f(z_1, \ldots, z_n)$ be analytic in $|z_1| < 1, \ldots, |z_n| < 1$ and satisfy*

$$\int_0^{2\pi} \cdots \int_0^{2\pi} |f(r_1 e^{i\theta_1}, \ldots, r_n e^{i\theta_n})|^\lambda d\theta_1 \ldots d\theta_n < C^\lambda;\ r_1, \ldots, r_n < 1;\ \lambda > 0, \tag{8}$$

then

[3]The reference at the end of **(4)** to ergodic theorems for compact groups is erroneous or at least misleading. The theorems actually meant are analogous to ergodic theorems **(6)** but deal with averages over sets tending to zero (like derivatives).

$$(9) \qquad \int_0^{2\pi} \cdots \int_0^{2\pi} \{\sup_{\Delta} |f(r_1 e^{i\theta_1}, \ldots, r_n e^{i\theta_n})|\}^{\lambda} d\theta_1 \ldots d\theta_n < \alpha_n C^{\lambda}$$

where α_n, C^{λ} have the same meanings as before and Δ is the region, for fixed $\theta_1, \ldots, \theta_n$, described by

$$(10) \qquad 0 < \frac{1 - r_i}{1 - r_j} < K; \qquad i, j = 1, \ldots, n, i \neq j,$$

K being any positive constant.

The necessity for (10) is related to the fact that the values of a function harmonic in each $|z_i| < 1$ are determined solely by its boundary values on T_n, which is thus a "distinguished boundary surface" in Bergman's terminology.

If in Theorem 3 and Theorem A one observes that when dealing with T_n, spheres may be replaced by hyper-cubes, while $C = 3^n$, then one will see that Theorem 4 follows from them as Theorem 3 does—provided, that is, that one proves the more delicate version of the connecting link between Theorems 1 and 3 (§4).

2. Proof of Theorem A. Consider any point P of S whose $\sigma(P)$ overlaps only a finite number of $\sigma(P)$. This certainly implies that we can find some sphere (not a $\sigma(P)$!) with P as center such that within this sphere there is no other point of S. This, by familiar reasoning, implies that the set of such points P is denumerable, hence of measure 0. Let us, therefore, discard this set and the $\sigma(P)$ belonging to it and ignore them in further reasoning.

Let B_1 be the least upper bound of the radii of $\sigma(P)$. Obviously one may assume $B_1 < \infty$. Otherwise, the theorem is trivial.

After this remark I shall, in fact, prove (7) with S replaced by the set consisting of the union σ of all $\sigma(P)$, and $m(S)$ replaced by $m(\sigma) - \epsilon$ for any ϵ. This, of course, will prove (7), since σ contains S.

Let $V(B_1)$ be the volume of a sphere of radius B_1. By the definition of B_1, one can choose $\sigma(P_1)$ such that its volume $V_1 > V(B_1) - \frac{1}{2}\epsilon$ (obviously $V_1 < V(B_1)$). Let σ_1 be the set consisting of the union of $\sigma(P_1)$ and all adjoining $\sigma(P)$. The volume of σ_1 is not greater than $CV(B_1)$. In fact, let $\bar{\sigma}$ be a sphere of radius B_1 about P_1 (not a $\sigma(P)$!) and hence of volume $V(B_1)$. Now σ_1 is certainly contained in the union of $\bar{\sigma}$ and all spheres adjoining it. But these latter are certainly of radius $<B_1$ by the definition of B_1. Hence Property A implies that the union in question has volume $<CV(B_1)$. Let $V(B_2)$ be the least upper bound of the volumes of those $\sigma(P)$ not in σ_1. Choose such a $\sigma(P_2)$ whose volume $V_2 > V(B_2) - \frac{1}{4}\epsilon$. Let σ_2 be the union of $\sigma(P_2)$ and all adjoining $\sigma(P)$ not in σ_1. As before, the volume of σ_2 is $<CV(B_2)$.

Continue this process inductively. One obtains a sequence $\sigma(P_k)$, obviously disjoint, with volumes V_k subject to these inequalities, where the $V(B_k)$ are defined similarly,

$$\sum_{k=1}^{\infty} \left(V(B_k) - \frac{\epsilon}{2^k} \right) < \sum_{k=1}^{\infty} V_k < m(\sigma) .$$

Since $\sum_k V(B_k)$ is convergent $V(B_k) \to 0$. This implies that the union of σ_k where σ_k is defined inductively as above, exhausts all $\sigma(P)$; for, if it did not, but omitted, say, one $\sigma(P')$, then from some k' on $V(B_k)$ would equal the volume of $\sigma(P')$.

Summing up, one has $\sigma = \Sigma\sigma_k$. Therefore,

$$\frac{1}{C} m(\sigma) < \Sigma\, V(B_k)$$

but

$$\Sigma\, V_k > \Sigma\, V(B_k) - \epsilon > \frac{1}{C} m(\sigma) - \epsilon.$$

Now one chooses K so that

$$\sum_{k=1}^{K} V_k > \sum_{1}^{\infty} V_k - \epsilon$$

and one has finally

$$\frac{1}{C} m(\sigma) - 2\epsilon < \sum_{k=1}^{K} V_k.$$

LEMMA 1. *Theorem* A *applies to* S_{n-1} *with spherical caps as the* $\sigma(P)$ *and to* T_n *with hypercubes,* $-\phi < \theta_i < \phi$, $(i = 1, \ldots, n)$, *as* $\sigma(P)$, *where* $C = 3^{n-1}$ *in the first and* $C = 3^n$ *in the second case.*

Proof. The first part is obvious as is the very last statement. In dealing with S_{n-1} I remark that

$$V_r = C_{n-1} \int_0^r \sin^{n-2}\theta\, d\theta.$$

Now the volume of a sphere of radius r plus those adjoining it of smaller r radius is certainly less than or equal to that of a sphere of radius $3r$. Since $3r < \pi$,

$$\sin 3r = 3 \sin r - 4 \sin^3 r < 3 \sin r$$

so that $\sin^{n-2} 3r < 3^{n-2} \sin^{n-1} r$. Therefore

$$\int_0^{3r} \sin^{n-2}\theta\, d\theta = 3 \int_0^r \sin^{n-2} 3\,\theta' d\theta' < 3^{n-1} \int_0^r \sin^{n-2}\theta d\theta\,.$$

3. Proof of Theorem 3. The key to Theorem 3 is the important

THEOREM 5. *The measure of the set* S_α *of points* P *for which* $f^*(P) > \alpha$ *does not exceed*

$$\frac{3^{n-1}}{\alpha} \int_{S_{n-1}} |f(P)|\, dV_P$$

It also does not exceed

$$\frac{2.3^{n-1}}{\alpha} \int_{|f(P)| > \frac{1}{2}\alpha} |f(P)|\, dV_P.$$

Proof. For each $P \in S_\alpha$ by definition one can find an r_P such that

$$\int_{\sigma P(r_P)} |f(P')| \, dV_{P'} > V(r_P)\alpha.$$

By Lemma 1 one can find a finite number of the $\sigma_P(r_P)$ whose total measure exceeds $3^{-(n-1)} m(S_\alpha) - \epsilon$. One has then

$$\int_{S_{n-1}} |f(P)| \, dV_P > \int_{\Sigma} |f(P)| \, dV_P > \frac{1}{3^{n-1}} m(S_\alpha) \, \alpha - \epsilon$$

where Σ is the finite set of $\sigma_P(r_P)$. The last statement is proved as follows: Let $h(P) = |f(P)|$ when $|f(P)| > \frac{1}{2}\alpha$, otherwise zero. Let $h^*(P)$ be defined in the same manner as $f^*(P)$. Obviously, we have $f^*(P) < h^*(P) + \frac{1}{2}\alpha$. Consequently $m(S_\alpha) <$ the measure of the set of P for which $h^*(P) > \frac{1}{2}\alpha$, which by the preceding part of the theorem is

$$< \frac{2.3^{n-1}}{\alpha} \int_{S_{n-1}} h(P) \, dV_P = \frac{2.3^{n-1}}{\alpha} \int_{|f(P)| > \frac{1}{2}\alpha} |f(P)| \, dV_P.$$

A similar proof yields

THEOREM 6. *Let $f(P)$ belong to L on the multitorus T_n. Let $\gamma(P)$ be the "cube"*

$$\theta_{i_P} - \phi < \theta_i < \theta_{i_P} + \phi$$

with P as center and side 2ϕ. Then the measure $m(S_\alpha)$ of the set of points P where

$$f^*(P) \equiv \underset{0<\phi<\pi}{\text{l.u.b.}} \frac{1}{2^n \phi^n} \int_{\gamma(P)} f(P') \, dV_{P'} > \alpha$$

is

$$< \frac{3^n}{\alpha} \int_{T_n} |f(P)| \, dV_P,$$

where dV_P is the volume element on T_n. It also does not exceed

$$\frac{2.3^n}{\alpha} \int_{|f(P)| > \frac{1}{2}\alpha} |f(P)| \, dV_P.$$

Proof of Theorem 3. Let $m(x)$ be the measure of the set of points where $|f(P)| > x$, and $m^*(x)$ be the measure of the set of points where $f^*(P) > x$. If $s(x)$ is any non-negative increasing function of x then

$$\int_{S_{n-1}} s(f(P)) \, dV_P = - \int_0^\infty s(x) \, dm(x)$$

$$\int_{S_{n-1}} s(f^*(P)) \, dV_P = - \int_0^\infty s(x) \, dm^*(x)$$

(7, p. 242).

Since, from Theorem 5,

$$m^*(x) < \frac{2.3^{n-1}}{x}\int_{|f(P)|>\frac{1}{2}x} |f(P)|\, dV_P = -\frac{2.3^{n-1}}{x}\int_{\frac{1}{2}x}^{\infty} y\, dm(y),$$

by formal substitution and interchange of integrations we have

$$\int_0^{\infty} m^*(x)\, x^{p-1}dx < -2.3^{n-1}\int_0^{\infty} x^{p-1}dx \int_{\frac{1}{2}x}^{\infty} y\, dm(y)$$

$$= -2.3^{n-1}\int_0^{\infty} y dm\,(y)\int_0^{2y} x^{p-2}dx$$

$$= \frac{-2^p.3^{n-1}}{(p-1)}\int_0^{\infty} y^p dm(y).$$

But this latter

$$= \frac{2^p.3^{n-1}}{(p-1)}\int_{S_{n-1}} |f(P)|^p dV_P$$

and is, therefore, finite. As a consequence

$$\lim_{\xi\to\infty}\int_{\xi}^{2\xi} m^*(x)\, x^{p-1}dx = 0;$$

however, since $m^*(x)$ is a decreasing function

$$m^*(2\xi)\,\xi^p\frac{2^p-1}{p} = m^*(2\xi)\int_{\xi}^{2\xi} x^{p-1}dx < \int_{\xi}^{2\xi} m^*(x)\, x^{p-1}dx;$$

therefore, $\lim m^*(\xi)\xi^p = 0$, and we can integrate by parts, getting

$$\int_{S_{n-1}} \{f^*(P)\}^p dV_p = -\int_0^{\infty} x^p dm^*(x) < \frac{2^p.3^{n-1}}{(p-1)}\int_{S_{n-1}} |f(P)|^p dV_P,$$

which is (5) with $C_{n,p} = 2^p.3^{n-1}/(p-1)$.

The second statement has been proved by Hardy-Littlewood **(2)** for $n = 2$. The third statement has a similar proof. This time we notice that for the same reasons

$$\int_1^{\infty} m^*(x)\, dx < -2.3^{n-1}\int_1^{\infty} y dm\,(y)\int_1^{2y}\frac{dx}{x} = -2.3^{n-1}\int_1^{\infty} y\log 2y\, dm(y)$$

$$< 2.3^{n-1}\int_{S_{n-1}} |f(P)|\log^+|f(P)|\, dV_P + 2.3^{n-1}\log 2\int_{S_{n-1}} |f(P)|\, dV_P\,.$$

Integrating by parts and noting that

$$|f| < e + |f|\log^+|f|$$

and

$$\int_0^1 m^*(x)dx < V(S_{n-1})$$

we have (6). These constants are not as good as those of Hardy-Littlewood in the original case, $n = 2$.

THEOREM 7. *Let $f(P)$ belong to L^p, $p > 1$ on T_n. Then if $f^*(P)$ is defined as in Theorem 7*

$$\int_{T_n} \{f^*(P)\}^p dV_P < C_{n,p} \int_{T_n} |f(P)|^p dV_P$$

where $C_{n,p}$ depends only on n and p. This is no longer true for $p = 1$; however, if $|f(P)| \log^+|f(P)|$ is integrable on T_n, then

$$\int_{T_n} f^*(P)\, dV_P < B_n \int_{T_n} |f(P)| \log^+ |f(P)|\, dV_P + C_n$$

where B_n depends only on n.

The proof is exactly like that of Theorem 3 with appropriate changes.

It will now be seen immediately that Theorem 3 can be extended to an arbitrary space with Property A, where 3^{n-1} is replaced by C.

4. Theorems which relate radial suprema to averages.

THEOREM 8. *Let $f(P)$ belong to L on S_{n-1}. Let $u(r, P)$ be the harmonic function in S_n which takes on the values $f(P)$ on S_{n-1}. If*

$$U(P) = \sup_{0<r<1} |u(r, P)|$$

then $U(P) < A_n f^*(P)$, where $f^*(P)$ is defined as in Theorem 3 and A_n is a constant depending only on n.

Proof. Define polar coordinates in n-space:

$$\begin{aligned}
x_1 &= r \cos \theta_1, \\
x_2 &= r \sin \theta_1 \cos \theta_2, \\
&\cdots \\
x_{n-1} &= r \sin \theta_1 \ldots \sin \theta_{n-2} \cos \theta_{n-1}, \\
x_n &= r \sin \theta_1 \ldots \qquad \sin \theta_{n-1}.
\end{aligned}$$

For fixed P we may assume that P is the point $x_1 = 1$, $x_i = 0$; $i > 1$, in which case θ_1 becomes the geodesic distance of any other point on S_{n-1} from P. We have then

$$u(r, P) = \frac{1}{\omega_n} \int_0^\pi d\,\theta_1 \ldots \int_0^\pi d\,\theta_{n-2} \int_{-\pi}^{\pi} P_n(r, \theta_1) \sin^{n-2}\theta_1 f(Q) \omega \cdot d\,\theta_{n-1}$$

where

$$\omega_n = \frac{2(\sqrt{\pi})^n}{\Gamma(\frac{1}{2}n)} \text{ and } P_n(r, \theta)$$

is the Poisson kernel for the sphere,

$$\frac{1 - r^2}{(1 - 2r \cos \theta + r^2)^{\frac{1}{2}n}}, \omega = \sin^{n-3}\theta_2 \ldots \sin \theta_{n-2},$$

and Q has coordinates $(\theta_1, \ldots, \theta_{n-1})$ (**1**, Chap. IV). Now let us observe that in order to prove the lemma, it is sufficient to prove $|u(r, P)| < A_n f^*(P)$ where A_n is fixed and independent of r. We integrate by parts, then, with respect to θ_1 and obtain

$$u(r, P) = \frac{1}{\omega_n}\left\{P_n(r, \pi)\int_0^\pi d\theta_1 \ldots \int_{-\pi}^{\pi} f(Q) \sin^{n-2}\theta_1\omega d\theta_{n-1}\right.$$

$$\left. - \int_0^\pi d\theta_1 \cdot \frac{dP_n(r, \theta_1)}{d\theta_1} \cdot \int_0^{\theta_1}\int_0^\pi d\theta_2 \ldots \int_{-\pi}^{\pi} d\theta_{n-1} f(Q') \sin^{n-2}\theta\omega d\theta\right\}$$

where the coordinates of Q' are $(\theta, \theta_2, \ldots, \theta_{n-1})$. Recalling the definition of $f^*(P)$, more explicitly

$$f^*(P) = \underset{0<\theta<\pi}{\text{l.u.b.}}\left(C_n \int_0^\theta \sin^{n-2}\theta\, d\theta\right)^{-1} \int_0^\theta \left[\int_0^\pi d\theta_2 \ldots \int_{-\pi}^{\pi} d\theta_{n-1} |f(Q)| \cdot \sin^{n-2}\theta \cdot \omega\right] d\theta$$

where

$$C_n = \int_0^\pi d\theta_2 \ldots \int_{-\pi}^{\pi} \omega d\theta_{n-1} = \omega_{n-1};$$

therefore

$$|u(r, P)| < K_n f^*(P) + C_n f^*(P) \int_0^\pi \left|\frac{dP_n}{d\theta_1}(r, \theta_1)\right| \cdot \left|\int_0^{\theta_1} \sin^{n-2}\theta\, d\theta\right| d\theta_1$$

$$< f^*(P)\left[K_n + C_n \int_{-\pi}^{\pi} \left|\frac{dP_n}{d\theta_1}(r, \theta_1)\right| \cdot |\theta_1| \cdot d\theta_1\right]$$

where K_n is a constant depending only on n. The last expression (**2**, p. 107) in brackets is $< D_n$. This completes the proof.

THEOREM 9. *Let $f(P)$ belong to L on T_n. Let $u(r_1, \ldots, r_n, P)$ be the function which is harmonic in the polycylinder,*

$$P_n\colon |z_i| < 1 \ (i = 1, \ldots, n)$$

and which assumes the values $f(P)$ on T_n. Then, if $f^*(P)$ is defined as in Theorem 7, we have

$$\sup_{(r_1, \ldots, r_n) \epsilon \Delta} |u(r_1, \ldots, r_n, P)| < A_\Delta f^*(P)$$

where Δ is the region of $0 < r_i < 1$, $(i = 1, \ldots, n)$, described by (10). *A_Δ depends only on Δ (i.e., K) and n.*

Proof. I remark that the latter restriction seems quite essential as is evidenced not only in the proof but by the implications of the lemma (cf. the next section). I observe again that one can assume that P is the point $\theta_i = 0$ $(i = 1, \ldots, n)$ on T_n. Because of a few complications it will be easier to present here the proof for $n = 2$ only. The extension to arbitrary n is straightforward.

First,

$$u(r_1, r_2, P) = \frac{1}{\pi 4^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} P(r_1, \theta_1)\, P(r_2, \theta_2)\, f(\theta_1, \theta_2)\, d\theta_1 d\theta_2$$

where $P(r, \theta) = (1 - r^2)/(1 - 2\cos\theta + r^2)$ is the usual Poisson kernel. Repeated integration by parts and interchange of integration gives, after taking absolute values

$$|u(r_1, r_2, P)| < \frac{1}{\pi 4^2}\left[\frac{1 - r_1}{1 + r_1}\cdot\frac{1 - r_2}{1 + r_2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} |f(\theta_1, \theta_2)|\, d\theta_1 d\theta_2\right.$$

$$+ \frac{1 - r_2}{1 + r_2} \int_{-\pi}^{\pi} \left|\frac{d}{d\theta_1} P(r_1, \theta_1)\right| d\theta_1 \int_0^{\theta_1} \int_{-\pi}^{\pi} |f(\theta, \theta')|\, d\theta d\theta'$$

$$+ \frac{1 - r_1}{1 + r_1} \int_{-\pi}^{\pi} \left|\frac{d}{d\theta_2} P(r_2, \theta_2)\right| d\theta_2 \int_0^{\theta_2} \int_{-\pi}^{\pi} |f(\theta, \theta')|\, d\theta d\theta'$$

$$+ \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \left\{\left|\frac{d}{d\theta_2} P(r_2, \theta_2)\, \frac{d}{d\theta_1} P(r_1, \theta_1)\right|\right.$$

$$\left.\left.(\operatorname{sgn}\theta_1 \cdot \operatorname{sgn}\theta_2) \int_0^{\theta_1} \int_0^{\theta_2} |f(\theta, \theta')|\, d\theta d\theta'\right\} d\theta_1 d\theta_2\right]$$

$$= \frac{1}{4\pi^2}[I_1 + I_2 + I_3 + I_4].$$

Recalling the definition of $f^*(P)$ one has $I_1 < 4\pi^2 f^*(P)$. To get an inequality for I_2 (and I_3) one observes that the inner double integral is less than or equal to

$$\int_{-\pi}^{\pi} \int_{-\pi}^{\pi} |f(\theta_1, \theta_2)|\, d\theta_1 d\theta_2 < 4\pi^2 f^*(P).$$

Furthermore,

$$\frac{1 - r_2}{1 + r_2} \int_{-\pi}^{\pi} \left|\frac{d}{d\theta_1} P(r_1, \theta_1)\right| d\theta_1$$

$$= \frac{1 - r_2}{1 + r_2}\left[- \int_0^{\pi} \frac{d}{d\theta_1} P(r_1, \theta_1)\, d\theta_1 + \int_{-\pi}^{0} \frac{d}{d\theta_1} P(r_1, \theta_1)\, d\theta_1\right]$$

$$= \frac{1 - r_2}{1 + r_2}\left[+ 2\left(\frac{1 + r_1}{1 - r_1}\right) - 2\left(\frac{1 - r_1}{1 + r_1}\right)\right] < 4K.$$

Therefore I_2 (and I_3) $< 4\pi^2 \,.\, 4Kf^*(P)$. Next, in I_4 the inner double integral is less than or equal to

$$\operatorname{sgn}\theta_1 \cdot \operatorname{sgn}\theta_2 \int_0^{\theta_2} \int_0^{(\operatorname{sgn}\theta_1)|\theta_2|} < \begin{cases} \displaystyle\int_{-\theta_2}^{\theta_2} \int_{-\theta_2}^{\theta_2}, & |\theta_1| < |\theta_2|, \\ \displaystyle\int_{-\theta_1}^{\theta_1} \int_{-\theta_1}^{\theta_1}, & |\theta_1| > |\theta_2|. \end{cases}$$

We split up I_4,

$$I_4 = \iint_{|\theta_1|<|\theta_2|} + \iint_{|\theta_1|>|\theta_2|}$$

which with the previous remark gives us

$$I_4 < 4f^*(P)\Bigg[\iint_{|\theta_1|>|\theta_2|} \frac{d}{d\theta_1} P(r_1, \theta_1) \cdot \frac{d}{d\theta_2} P(r_2, \theta_2)| \cdot |\theta_1|^2 d\theta_1 d\theta_2$$

$$+ \iint_{|\theta_1|<|\theta_2|} \left|\frac{d}{d\theta_1} P(r_1, \theta_1) \cdot \frac{d}{d\theta_2} P(r_2, \theta_2)\right| \cdot |\theta_2|^2 d\theta_1 d\theta_2\Bigg],$$

the first integral in brackets is less than

$$2\pi r_1(1 - r_1^2) \int_{-\pi}^{\pi} \left|\frac{\theta \sin \theta}{(1 - 2r_1 \cos \theta + r_1^2)^2}\right| d\theta \cdot \int_{-\pi}^{\pi} \left|\frac{d}{d\theta} P(r_2, \theta)\right| d\theta$$

$$< 2r_1\pi(1 + r_1)(1 - r_1)\left[2\,\frac{(1 + r_2)}{1 - r_2}\right] \cdot \int_{-\pi}^{\pi} \left|\frac{\theta \sin \theta}{(1 - 2r_1 \cos \theta + r_1^2)^2}\right| d\theta.$$

The integral on the right is less than or equal to a constant C (the reasoning is the same as in the Hardy-Littlewood reference in the proof of Lemma 3). Therefore, this first integral in brackets (and similarly the other) is less than or equal to $16\pi K\,C$. Setting

$$A_\Delta = (1 + 8K + 32KC/\pi),$$

one completes the proof.

5. Proofs of Theorems 1, 2, and 4 and related theorems.

THEOREM 10. *Let $f(P)$ belong to L^p, $p > 1$ on S_{n-1}. Let $u(r, P)$ be the function harmonic in S_n with boundary values $f(P)$. If $U(P) = \sup\limits_{0<r<1} |u(r, P)|$, then*

$$\frac{1}{V_1} \int_{S_{n-1}} \{U(P)\}^p dV_P < C_{n,p} \frac{1}{V_1} \int_{S_{n-1}} |f(P)|^p dV_P$$

where $C_{n,p}$ is a constant, depending only on n and p. For $p = 1$ this is not true. However, if $|f(P)| \log^+|f(P)|$ is integrable on S_{n-1}, then

$$\frac{1}{V_1} \int_{S_{n-1}} U(P)\, dV_P < \frac{B_n}{V_1} \int_{S_{n-1}} |f(P)| \log^+|f(P)|\, dV_P + \frac{C_n}{V_1}$$

where B_n and C_n depend only on n.

Proof. The first and third statements are corollaries of Theorem 3 and Theorem 9. The second statement has been proved in **(2)** for $n = 2$. As a corollary of Theorem 10 we have

THEOREM 11. *Let $u(r, P)$ be harmonic in S_n and let it be such that*

$$\text{(11)} \qquad \frac{1}{V_1} \int_{S_{n-1}} |u(r, P)|^p dV_P < C^p$$

for all $r < 1$, and fixed $p > 1$. Then if $U(P)$ is defined as in Theorem 10

$$\frac{1}{V_1}\int_{S_{n-1}}\{U(P)\}^p dV_P < C_{n,p}C^p \tag{12}$$

Proof. Let

$$U_R(P) = \sup_{0<r<R<1} |u(r, P)|;$$

then

$$\frac{1}{V_1}\int_{S_{n-1}}\{U_R(P)\}^p dV_P < \frac{C_{n,p}}{V_1}\int_{S_{n-1}} |u(R, P)|^p dV_P < C_{n,p}C^p,$$

by Theorem 10. Now $U_R(P) \uparrow U(P)$ as $R \to 1$. Hence, Lebesgue's monotone convergence theorem completes the proof.

THEOREM 12. *Let $u(r, P)$ satisfy* (11). *Then there exists a function $u(P) \in L^p$ on S_{n-1} such that*

$$\lim_{r\to 1}\int |u(r, P) - u(P)|^p dV_P = 0 \tag{13}$$

Proof. Hölder's inequality implies

$$\frac{1}{V_1}\int_{S_{n-1}} |u(r, P)|\, dV_P < \left(\frac{1}{V_1}\int_{S_{n-1}} |u(r, P)|^p dV_P\right)^{1/p} < (C_{n,p})^{1/p}.$$

As a result

$$F(r, S) = \int_S u(r, P)\, dV_P,$$

where S is a measurable subset of S_{n-1}, constitute a set of absolutely continuous set-functions on S_{n-1} which are uniformly bounded. According to Radon's theory of integration one can form, for any Φ continuous on S_{n-1}, the Radon-Stieltjes integral

$$\int_{S_{n-1}} \Phi\, dF(r, S).$$

Now it is a classic theorem of Radon that from the uniformly bounded set of set-functions $F(r, S)$ on S_{n-1} one can extract a sequence $F(r_m, S)$ and find another bounded set-function $F(S)$ such that, for any continuous Φ,

$$\lim_{m\to\infty}\int_{S_{n-1}} \Phi\, dF(r_m, S) = \int_{S_{n-1}} \Phi\, dF(S) \tag{14}$$

where the $r_m \to 1$, otherwise the theorem is trivial.
Now

$$|F(r, S)| < \int_S |u(r, P)|\, dV_P < \int_S U(P)\, dV_P \tag{15}$$

so that the $F(r, S)$ are uniformly absolutely continuous, since $U(P)$ by (12) and Holder's inequality belongs to L. Therefore, by choosing Φ in (14) to be

the characteristic function (rounded-off) of S one sees that $F(S)$ is also absolutely continuous, and, therefore, the integral of a point function $u(P)$ of class L. Accordingly, if one picks Φ in (14) to be the Poisson kernel one finds

$$u(r, P) = \int_{S_{n-1}} P_n(r, Q)\, u(Q)\, dV_Q. \tag{16}$$

One also sees from (15), by applying Lebesgue's differentiation theorem, that $|u(r, P)| < U(P)$ almost everywhere so that $u(P)$ is in L^p by (12).

The reasoning in (**7**, p. 85) using (16) completes the proof of (13).

To prove (2), I observe first that a subharmonic function $w(r, p)$ satisfying (11) also satisfies the analogue of (12). The proof of this is reduced to (12) by the device of the harmonic majorant of $w(r, P)$ and is to be found in (**2**, p. 113, footnote 1) which carries over word for word to several variables.

Next, following Hardy and Littlewood, I set $w(r, P) = |f|^{\frac{1}{2}\lambda}$ and observe that (1) implies that $w(r, P)$ satisfies (11) with $p = 2 > 1$. That $|f|^{\frac{1}{2}\lambda}$ is subharmonic is a simple consequence of the mean-value theorem for the function $f^{\frac{1}{2}\lambda}$ which is analytic in the neighborhood of any point where $f \neq 0$. Then (2) follows immediately.

The proof of (9) follows from Theorems 6 and 9 through the intermediary of theorems analogous to 10 and 11 exactly as in the proof of (2).

Finally (3), Theorem 2, follows from Theorem 12 when $\lambda > 1$ since $|Rf| < |f|$ and $|If| < |f|$ and the convergence of f follows from that of Rf and If by Minkowski's inequality. When $\lambda = 1$ the Hardy-Littlewood inequality is still valid for f, unlike the harmonic function; therefore the reason, ing of Theorem 12 may be repeated to account for this case.

Obviously, a similar theorem may be deduced for the polycylinder (cf. also **10**).

References

1. R. Courant and D. Hilbert, *Methoden der Mathematischen Physik*, vol. II (Berlin, 1937)
2. G. H. Hardy and J. E. Littlewood, *A maximal theorem with function-theoretic applications*, Acta Math., *54* (1930), 81–116.
3. H. E. Rauch. *Generalisation d'une proposition de Hardy et Littlewood et de théorèmes ergodiques qui s'y rattachent*, Comptes Rendus, *227* (1948), 887–889.
4. ———, *A Poisson formula and the Hardy and Littlewood theorem for matrix spaces*, Bull. Amer. Math. Soc., *55* (1949), Abstract 250.
5. K. T. Smith, *A generalization of an inequality of Hardy and Littlewood*, Can. J. Math., *8* (1956), 157-170.
6. N. Wiener, *The ergodic theorem*, Duke Math. J., *5* (1939), 1–18.
7. A. Zygmund, *Trigonometrical Series* (Warsaw, 1935).
8. ———, *On the existence of boundary values for regular functions of several complex variables*, Bull. Amer. Math. Soc., *54* (1948), Abstract 502t.
9. ———, *A remark on functions of several complex variables*, Acta Szeged, *12* (1950), 66–68.
10. ———, *On the boundary values of functions of several complex variables*, Fund. Math., *36* (1949), 207–305.

University of Pennsylvania

AN EXTENSION PROBLEM FOR FUNCTIONS WITH MONOTONIC DERIVATIVES

W. B. JURKAT

Introduction. This paper deals with questions of the following type. Problem (A): Let $F(x)$ be the nth integral of a positive non-decreasing function for all large positive x, the problem is to find a function $f(x)$, being the nth integral of a non-decreasing function for all x $(-\infty < x < \infty)$, with the property

$$\text{(A)} \qquad f(x) = \begin{cases} F(x), & \text{for all large positive } x; \\ 0, & \text{for all large negative } x. \end{cases}$$

Problem (A) can be considered as a special case of the boundary value problems, which we discuss in §2. Roughly speaking, the question is here what values may be assumed by the nth integral of a monotonic function and its first n derivatives at the boundary of an interval. It is no loss of generality to suppose the left-hand boundary values to be equal to zero as can be seen by subtracting a suitable polynomial. Then the solution of the problem directly depends on the solution of the reduced Hausdorff and Stieltjes moment problems, for the latter of which we give a new approach (§1).

The method indicated leads in a simple manner to a complete solution of problem (A), depending on the behaviour of certain quadratic forms (§3). The main result of the paper consists of determining this behaviour for a large class of functions $F(x)$, for which therefore problem (A) can be settled (Theorem 6).

1. Some reduced moment problems. If a finite sequence μ_ν $(\nu = 0, \ldots, n)$ is given, the reduced Hausdorff moment problem is to determine a non-decreasing function $\psi(t)$, $0 < t < 1$, such that

$$\text{(I)} \qquad \mu_\nu = \int_0^1 t^\nu d\psi(t) \qquad (\nu = 0, \ldots, n).$$

The following result is essentially due to Achyeser-Krein (**1**; see also **3**; **5**, pp. 29–30; **6**, p. 77).

THEOREM 1. *A necessary and sufficient condition that the moment problem* (I) *should have a solution is that, in case $n = 2m$, both quadratic forms*

$$\text{(1)} \qquad \sum_{i,j=0}^{m} \mu_{i+j} x_i x_j, \qquad \sum_{i,j=0}^{m-1} (\mu_{i+j+1} - \mu_{i+j+2}) x_i x_j$$

Received March 4, 1955. This research was supported by the United States Air Force, through the Office of Scientific Research of the Air Research and Development Command, under Contract Number AF 18 (600)–691.

should be non-negative, whereas in case $n = 2m + 1$, both quadratic forms

$$\sum_{i,j=0}^{m} \mu_{i+j+1} x_i x_j, \qquad \sum_{i,j=0}^{m} (\mu_{i+j} - \mu_{i+j+1}) x_i x_j \tag{2}$$

should be non-negative.

If a finite sequence $\mu_\nu (\nu = 0, \ldots, n)$ is given, we also consider the reduced moment problem of determining a non-decreasing function $\psi(t)$, $0 < t < T$, such that

$$\mu_\nu = \int_0^T t^\nu d\psi(t) \tag{II}$$

with suitable $T > 0$.

If (II) has a solution for a certain $T = T_0$, then it also has for all $T > T_0$. Replacing μ_ν by μ_ν / T^ν and x_i / T^i by x_i in Theorem 1 we obtain at once

THEOREM 2. *A necessary and sufficient condition that the moment problem* (II) *should have a solution is that for $n = 2m$ or $n = 2m + 1$ respectively*

$$\sum_{i,j=0}^{m} \mu_{i+j} x_i x_j > 0, \qquad \sum_{i,j=0}^{m-1} \mu_{i+j+2} x_i x_j < T \sum_{i,j=0}^{m-1} \mu_{i+j+1} x_i x_j \tag{3}$$

or

$$\sum_{i,j=0}^{m} \mu_{i+j+1} x_i x_j > 0, \qquad \sum_{i,j=0}^{m} \mu_{i+j+1} x_i x_j < T \sum_{i,j=0}^{m} \mu_{i+j} x_i x_j \tag{4}$$

should hold for all values of x_i and a suitable $T > 0$.

COROLLARY. *A sufficient condition is that for $n = 2m$ or $n = 2m + 1$ respectively the quadratic forms*

$$\sum_{i,j=0}^{m} \mu_{i+j} x_i x_j, \qquad \sum_{i,j=0}^{m-1} \mu_{i+j+1} x_i x_j \tag{5}$$

or

$$\sum_{i,j=0}^{m} \mu_{i+j+1} x_i x_j, \qquad \sum_{i,j=0}^{m} \mu_{i+j} x_i x_j \tag{6}$$

should be positive definite, whereas a necessary condition is that they should be non-negative.

It is worthwhile to point out the close connection between problem (II) and the reduced Stieltjes moment problem of determining a non-decreasing function $\psi(t)$, $0 < t < \infty$, such that

$$\mu_\nu = \int_0^\infty t^\nu d\psi(t) \qquad (\nu = 0, \ldots, n). \tag{III}$$

Necessary and sufficient conditions for the moment problem (III) to have a solution are due to Verblunsky **(7)**, who based his argument on certain algebraic lemmas of E. Fischer. By means of Theorem 1 or 2 we are able to give a new and very simple approach, avoiding with Verblunsky the theory of

continued fractions. This approach will also give us more detailed information about the solution.

First we infer from Theorem 13b of Widder (**8**, p. 138) that *a necessary and sufficient condition that the moment problem* (III) *should have a solution* $\psi(t)$ *with infinitely many points of increase is that the quadratic forms* (5) *resp.* (6) *should be positive definite* (cf **8**, p. 6). The necessary part is trivial and the sufficient part follows by inductive definition of $\mu_{n+1}, \mu_{n+2}, \ldots$.

Now the Corollary to Theorem 2 shows that, if there is a solution of (III) with infinitely many points of increase, then there also is a solution of (II). If there is a solution of (III) with a finite number of points of increase (that is, a step-function with finitely many steps) the same conclusion is true. Conversely every solution of (II) also gives a solution of (III), such that *the problems* (II) *and* (III) *are equivalent.* The conditions of Theorem 2, now also valid for problem (III), are slightly different from those of Verblunsky and have the advantage of using only the known values $\mu_0, \ldots, \mu_n$.

With the help of a mean value theorem for systems of integrals (**4**, p. 97) we further see that, if there is a solution of (II), then there also is a solution of (II) by a non-decreasing step-function with finitely many steps. Hence, if (III) has a solution, then (III) even has a solution with a finite number of points of increase. Therefore *the conditions of Theorem* 2 *also are necessary and sufficient for the moment problem* (III) *to have a solution* $\psi(t)$ *with a finite number of points of increase.*

Similar arguments can be used for the reduced Hamburger moment problem

$$\text{(IV)} \qquad \mu_\nu = \int_{-\infty}^{\infty} t^\nu d\psi(t) \qquad (\nu = 0, \ldots, n).$$

2. Some boundary value problems for functions with monotonic derivatives. We consider the following problem (B): Given real numbers $c_\nu (\nu = 0, \ldots, n)$, X, and $T > 0$, find a function $f(x)$, which is for $X - T < x < X$ the nth integral of a non-decreasing function $\phi(x)$ and satisfies the boundary value condition

$$\text{(3)} \qquad f^{(\nu)}(X - T) = 0, \qquad f^{(\nu)}(X) = c_\nu \qquad (\nu = 0, \ldots, n),$$

where $f^{(n)}(x)$ is to be identified with $\phi(x)$ by definition. (For points of continuity $f^{(n)}(x) = \phi(x)$ holds by itself, and for other points $f^{(n)}(x)$, $n > 1$, is not defined a priori.) Besides (B) we introduce the problem (B′) differing from (B) only in the possibility that $T > 0$ may be chosen suitably.

If problem (B) or (B′) has a solution, then for $X - T < x < T$ and $\nu = 0, \ldots, n$ we have

$$\text{(7)} \qquad f^{(\nu)}(x) = \frac{1}{(n-\nu)!} \int_{X-T}^{x} (x - t)^{n-\nu} d\phi(t);$$

in particular,

$$\text{(8)} \qquad c_\nu = \frac{1}{(n-\nu)!} \int_{X-T}^{X} (X - t)^{n-\nu} d\phi(t).$$

Conversely, if (8) holds with a non-decreasing function $\phi(x)$, where we may assume $\phi(X - T) = 0$ without restriction, then

$$f(x) = \frac{1}{n!}\int_{X-T}^{x} (x - t)^n d\phi(t) \tag{9}$$

is a solution of (B) or (B′).

By a change of variables the condition (8) can be written in the form

$$\nu!\, c_{n-\nu} = \int_0^T t^\nu d\psi(t) \qquad (\nu = 0, \ldots, n), \tag{10}$$

or

$$\nu!\, c_{n-\nu}/T^\nu = \int_0^1 t^\nu d\psi(t) \qquad (\nu = 0, \ldots, n), \tag{11}$$

with non-decreasing functions $\psi(t)$.

Thus we have proved the following results:

THEOREM 3. *Problem* (B) *has a solution if and only if problem* (I) *has a solution with*

$$\mu_\nu = \nu!\, c_{n-\nu}/T^\nu \qquad (\nu = 0, \ldots, n).$$

THEOREM 4. *Problem* (B′) *has a solution if and only if problem* (II) *has a solution with*

$$\mu_\nu = \nu!\, c_{n-\nu} \qquad (\nu = 0, \ldots, n).$$

Explicit conditions can be taken from Theorems 1 and 2.

3. Extension theorems for functions with monotonic derivatives. We now consider problem (A) at the beginning of this paper. A simple argument shows:

THEOREM 5. *Problem* (A) *has a solution if and only if problem* (B′) *has a solution with*

$$c_\nu = F^{(\nu)}(X) \qquad (\nu = 0, \ldots, n)$$

for all large positive X (X *only denoting numbers, where* $F^{(n)}(X)$ *exists*).

Explicit conditions can be taken from Theorem 2 by means of Theorem 4. We shall only use the special conditions of the Corollary of Theorem 2:

COROLLARY. *A sufficient* (*necessary*) *condition that problem* (A) *should have a solution is that for* $n = 2m$ *or* $n = 2m + 1$ *respectively the quadratic forms*

$$\sum_{i,j=0}^{m} (i+j)!\, F^{(n-i-j)}(X)\, x_i x_j, \qquad \sum_{i,j=0}^{m-1} (i+j+1)!\, F^{(n-i-j-1)}(X)\, x_i x_j \tag{12}$$

or

$$\sum_{i,j=0}^{m} (i+j+1)!\, F^{(n-i-j-1)}(X)\, x_i x_j, \qquad \sum_{i,j=0}^{m} (i+j)!\, F^{(n-i-j)}(X)\, x_i x_j \tag{13}$$

should be positive definite (non-negative) for all large positive X (X only denoting numbers, where $F^{(n)}(X)$ exists).

In general it is rather difficult to decide whether the forms (12), (13) are positive definite or not. But there is a certain class of functions $F(x)$ for which we can give a complete solution. These are the L-functions or logarithmico-exponential functions in the sense of Hardy (**2**, p. 17).

THEOREM 6. *Necessary and sufficient that the problem* (A) *should have a solution for a L-function $F(x)$ and $n \geqslant 2$ is that $F(x)/x^n$ should be non-decreasing for all large positive x. For $n = 0, 1$ there is always a solution.*

Proof. The existence of a solution for $n = 0, 1$ follows at once from the Corollary and the fact that $F^{(n)}(X)$, $F^{(n-1)}(X)$ are positive for $X \to +\infty$. From now on we may assume $n \geqslant 2$. Using the elementary properties of L-functions and our supposition on $F(x)$,

$$F^{(n+1)}(x) \geqslant 0, \quad F^{(n)}(x) \geqslant \epsilon > 0, \qquad x \to +\infty, \tag{14}$$

we only have to discuss the following cases:

$$xF^{(n+1)}(x) \geqslant \delta > 0, \qquad x \to +\infty \tag{a}$$

or

$$F(x) = \frac{x^n}{n!} L(x), \qquad 0 < \delta \leqslant L(x) = o(\log x), \qquad x \to +\infty, \tag{b}$$

with the possibilities, for $x \to +\infty$,

(b$_1$) $L'(x) = 0$, (b$_2$) $L'(x) > 0$, (b$_3$) $L'(x) < 0$.

We shall show that (A) has a solution in cases (a), (b$_1$), (b$_2$), where $F(x)/x^n$ is non-decreasing for $x \to +\infty$, and that (A) has no solution in case (b$_3$), where $F(x)/x^n$ is strictly decreasing for $x \to +\infty$. All this together will prove the Theorem.

Example. If $F(x) = x^n + x^{n-1}$, problem (A) has a solution for $n = 1$, but no solution for $n \geqslant 2$.

In case (a) we use the sufficient part of the Corollary and restrict ourselves to the form

$$\sum_{i,j=0}^{m} \frac{(i+j)!}{X^{i+j}} F^{(n-i-j)}(X)\, x_i x_j, \tag{15}$$

where $x_i X^i$ is replaced by x_i. We assume that (15) has its minimum value $M(X)$ on $\Sigma x_i^2 = 1$ for $x_i = x_i(X)$. It is enough to show that $M(X) > 0$ holds for $X \to +\infty$, or for any sequence $X = X_k \to \infty$ such that $x_i(X_k) \to \xi_i$ for $i = 0, \ldots, m$, with $\Sigma \xi_i^2 = 1$.

For $\nu = 0, \ldots, n$ and $X \geqslant x_0$ (x_0 large enough) we have

$$F^{(n-\nu)}(X) = \frac{1}{\nu!}\int_{x_0}^{X} (X-x)^{\nu} F^{(n+1)}(x)\,dx + \sum_{\zeta=0}^{\nu} \frac{(X-x_0)^{\zeta}}{\zeta!} F^{(n-\nu+\zeta)}(x_0) \tag{16}$$

and therefore

$$M(X) = \int_{x_0}^{X} \left(\sum_{i=0}^{m} \frac{(X - x)^i}{X^i} x_i \right)^2 F^{(n+1)}(x)\, dx$$
$$+ F^{(n)}(x_0) \left(\sum_{i=0}^{m} \frac{(X - x_0)^i}{X^i} x_i \right)^2 + O(X^{-1}).$$

Hence, taking $X = X_k$ and using (a),

$$M(X) \geqslant \frac{\delta}{X} \int_{x_0}^{X} \left(\sum_{i=0}^{m} \frac{(X - x)^i}{X^i} x_i \right)^2 dx + O(X^{-1})$$
$$\geqslant \frac{\delta}{X} \sum_{i,j=0}^{m} \frac{(X - x_0)^{i+j+1}}{i + j + 1} \frac{x_i x_j}{X^{i+j}} + O(X^{-1})$$
$$\geqslant \delta \sum_{i,j=0}^{m} \frac{\xi_i \xi_j}{i + j + 1} + o(1)$$
$$\geqslant \delta \int_0^1 \left(\sum_{i=0}^{m} \xi_i x^i \right)^2 dx + o(1).$$

Since

$$\int_0^1 \left(\sum_{i=0}^{m} \xi_i x^i \right)^2 dx > 0, \qquad \sum \xi_i^2 = 1$$

we obtain $M(X_k) > 0$ for $X_k \to +\infty$.

The same method can be used to show that the other forms in the Corollary are positive definite also.

In case (b_1), $F(x) = cx^n$ holds for $x \to +\infty$ with positive c. Then a solution of the problem (A) is given by the function

$$f(x) = \begin{cases} cx^n, & x \geqslant 0, \\ 0, & x < 0. \end{cases}$$

In case (b_2), instead of (16) we use the Leibniz formula

$$\text{(17)} \qquad \frac{\nu!}{X^\nu} F^{(n-\nu)}(X) = L(X) + \sum_{\zeta=1}^{n-\nu} \binom{n-\nu}{\zeta} L^{(\zeta)}(X) \frac{X^{\nu+\zeta}}{(\nu+\zeta)!} \frac{\nu!}{X^\nu}$$

for $\nu = 0, \ldots, n$ and $X \to +\infty$, and the asymptotic formulae

$$\text{(18)} \qquad XL'(X) = o(L(X))$$

and

$$\text{(19)} \qquad X^\zeta L^{(\zeta)}(X) = (-1)^{\zeta-1}(\zeta - 1)!\, XL'(X) + o(XL'(X))$$

for $\zeta = 1, 2, \ldots$ and $X \to +\infty$, which are consequences of (b) and Hardy (**2**, p. 37, line 5). From both together it follows for $\nu = 0, \ldots, n$ and $X \to +\infty$ that

$$\text{(20)} \qquad \frac{\nu!}{X^\nu} F^{(n-\nu)}(X) = L(X) + \gamma_\nu XL'(X) + o(XL'(X)),$$

where

$$(21)\qquad \begin{cases} \gamma_\nu = \sum_{\zeta=1}^{n-\nu} \binom{n-\nu}{\zeta} (-1)^{\zeta-1} \dfrac{\nu!(\zeta-1)!}{(\nu+\zeta)!} \\ \quad = \sum_{\zeta=1}^{n-\nu} \binom{n-\nu}{\zeta} (-1)^{\zeta-1} \int_0^1 (1-x)^\nu x^{\zeta-1} dx \\ \quad = \int_0^1 \dfrac{(1-x)^\nu - (1-x)^n}{x} dx, \qquad \nu = 0, \ldots, n. \end{cases}$$

Hence

$$(22)\qquad \begin{aligned} M(X) &= L(X)\left(\sum_{\iota=0}^{m} x_\iota\right)^2 + o(XL'(X)) \\ &\quad + XL'(X) \int_0^1 \left\{\left(\sum_{\iota=0}^{m} x_\iota(1-x)^\iota\right)^2 - (1-x)^n \left(\sum_{\iota=0}^{m} x_\iota\right)^2\right\} \frac{dx}{x}. \end{aligned}$$

Now let

$$X = X_k, \qquad \sum_{\iota=0}^{m} \xi_\iota \neq 0.$$

Then

$$M(X) = L(X)\left(\sum_{\iota=0}^{m} \xi_\iota\right)^2 + o(L(X)),$$

and on account of (b) we get $M(X_k) > 0$ for $X_k \to +\infty$. In the remaining case

$$\sum_{\iota=0}^{m} \xi_\iota = 0 \qquad (m > 1)$$

we have for $X = X_k$

$$(23)\qquad M(X) > XL'(X) \int_0^1 \left(\sum_{\iota=0}^{m} \xi_\iota (1-x)^\iota\right)^2 \frac{dx}{x} + o(XL'(X)).$$

Since

$$\int_0^1 \left(\sum_{\iota=0}^{m} \xi_\iota (1-x)^\iota\right)^2 \frac{dx}{x} > 0, \quad \sum_{\iota=0}^{m} \xi_\iota = 0, \quad \sum_{\iota=0}^{m} \xi_\iota^2 = 1,$$

we find $M(X_k) > 0$ for $X_k \to +\infty$. The same method can be used to show that the other forms in the Corollary are positive definite also.

In case (b_3) we use the necessary part of the Corollary and show that the minimum $M(X)$ of the form (15) on $\Sigma x_\iota^2 = 1$ is negative for $X \to +\infty$. Taking ξ_ι with

$$\sum_{\iota=0}^{m} \xi_\iota = 0, \quad \sum_{\iota=0}^{m} \xi_\iota^2 = 1 \qquad (m > 1),$$

we obtain similarly to (22) and (23)

$$
\begin{aligned}
M(X) &< \sum_{i,j=0}^{m} \frac{(i+j)!}{X^{i+j}} F^{(n-i-j)}(X)\, \xi_i \xi_j \\
&< XL'(X) \int_0^1 \left(\sum_{i=0}^{m} \xi_i \,(1-x)^i \right)^2 \frac{dx}{x} + o(XL'(X)) \\
&< 0, \qquad\qquad X \to +\infty.
\end{aligned}
$$

This completes the proof of Theorem 6.

References

1. N. Achyeser and M. Krein. *Über eine Transformation der reellen Toeplitzschen Formen und das Momentenproblem in einem endlichen Intervalle*, Commun. Soc. Math. Kharkoff (4), *11* (1935), 21–26.
2. G. H. Hardy, *Orders of infinity* (Cambridge Tract, 1924).
3. L. Kantorovič, *On the moment problem for a finite interval*, C. R. (Doklady), Acad. Sci. URSS (N.S.), *14* (1937), 531–537.
4. G. Kowalewski, *Integralgleichungen* (Berlin, 1930).
5. M. Krein, *The ideas of P. L. Čebyšev and A. A. Markov in the theory of limiting values of integrals and their further development*, Upsekhi Matem. Nauk (N.S.) *6* (1951), 3–120.
6. J. A. Shohat and J. D. Tamarkin, *The problems of moments* (New York, 1943).
7. S. Verblunsky, *On a problem of moments*, Proc. Cambridge Philos. Soc. *45* (1949), 1–4.
8. D. V. Widder, *The Laplace transform* (Princeton, 1946).

University of Cincinnati

ON RELATIONSHIPS AMONGST CERTAIN SPACES OF SEQUENCES IN AN ARBITRARY BANACH SPACE

C. W. McARTHUR

1. Introduction. Let X be a Banach space (B-space). A sequence $\{s(i)\}$ in X is *unconditionally summable* if and only if every rearrangement of the series $\sum_i s(i)$ is convergent. The set of unconditionally summable sequences in X will be written as $U(X)$. In this paper several classes of summable sequences in X will be compared with one another. Each class to be considered is identical with $U(X)$ when X has finite dimension.

The following notation will be used. The set of natural numbers will be denoted by N and the collection of non-null finite subsets of N by $\mathscr{F}$. A sequence in X will usually be denoted by the single letter s and its value at $i \in N$ by $s(i)$. If s is a sequence in X and $F \in \mathscr{F}$ the sum of the terms $s(i)$ such that $i \in F$ will be written $\sum_F s(i)$.

A sequence s in X will be called *weakly unconditionally summable* if and only if $\sum_i |f(s(i))| < \infty$ for every $f \in X^*$, the adjoint space of X. Let $B(X)$ stand for the set of weakly unconditionally summable sequences in X. Gelfand **(4)** has shown that $s \in B(X)$ if and only if $\sup[\|\sum_F s(i)\|: F \in \mathscr{F}] < \infty$. With the usual definitions for addition of sequences and multiplication of a sequence by a scalar $B(X)$ is a vector space. It is known that $B(X)$ is a B-space with the norm of each $s \in B(X)$ defined by $\|s\| = \sup[\|\sum_F s(i)\|: F \in \mathscr{F}]$. This will be the norm intended when $B(X)$ is referred to as a B-space in the sequel. As a consequence of a result of Birkhoff **(2)**, $U(X)$ is a closed linear subspace of $B(X)$.

Following Hadwiger **(5)**, a sequence s in a B-space X has an *invariant sum* if and only if there is an $x \in X$ such that $x = \sum_i s(i)$ and such that x is the sum of each of the convergent rearrangements of $\sum_i s(i)$. Let $IS(X)$ stand for the class of sequences in X with an invariant sum. It is known that if X has finite dimension then $U(X) = IS(X)$. Hadwiger **(5)** has shown that if X is a Hilbert space with infinite dimension then $U(X)$ is a proper subset of $IS(X)$. In this paper Hadwiger's result is sharpened and extended to any B-space with infinite dimension.

If s is a sequence in X and there is $x \in X$ such that $x = \sum_i s(i)$ then x will be called the *sum* of s. In case there is $x \sum X$ such that $f(x) = \sum_i f(s(i))$ for all $f \in X^*$ then x will be called the *weak sum* of s. It follows easily that *a sequence s in a B-space X can have at most one weak sum.* It can be shown that in any B-space X there are sequences which have a sum but are not elements of $B(X)$. Conversely, in some B-spaces, for example, in $X = c_0$, the B-space of real sequences which converge to 0 with $\|s\| = \sup[|s(i)|: i \in N]$ for each

$s \in c_0$, there exist sequences which are elements of $B(X)$ but which do not have sums.

Two new closed linear subspaces of $B(X)$ are introduced in this paper. They are

$B_w(X) = [s \in B(X): s$ has a weak sum$]$, $B_s(X) = [s \in B(X): s$ has a sum$]$.

For any B-space it is true that

$$U(X) \subset B_s(X) = IS(X) \cap B(X) \subset B_w(X) \subset B(X).$$

We show that if $X = c_0$ then all of these containments are proper.

2. Closed linear subspaces of $B(X)$**.** Dunford (3) and Gelfand (4) have shown that a sequence s in a B-space X is weakly unconditionally summable if and only if there is a real number M such that $\sum_i|f(s(i))| < M\|f\|$ for all $f \in X^*$. A norm for the vector space of weakly unconditionally summable sequences in X is defined by setting

$$\|s\|_1 = \sup[\textstyle\sum_i|f(s(i))|: f \in X^* \text{ and } \|f\| < 1]$$

for each sequence s of this class. Let $B'(X)$ denote the normed vector space of weakly unconditionally summable sequences in X with the norm of the preceding sentence. As a special case of a result of Dunford (3, Theorem 30) we have that $B'(X)$ is a B-space.

The following lemma is essentially given by Pettis (6, Theorem 3.2.2.).

Lemma 2.1. *If s is weakly unconditionally summable then*

$$\sup[\|\textstyle\sum_F s(i)\|: F \in \mathscr{F}] < \sup[\sum_i|f(s(i))|: f \in X^* \text{ and } \|f\| < 1]$$
$$< 2\sup[\|\textstyle\sum_F s(i)\|: F \in \mathscr{F}].$$

Lemma 2.2. *The normed vector space $B(X)$ is complete.*

Proof. Since $B(X)$ and $B'(X)$ differ only in their norms and $B'(X)$ is complete it is evident from the relationships between their norms given in Lemma 2.1 that $B(X)$ is complete.

Theorem 2.3. *For any B-space X the spaces $B_w(X)$ and $B_s(X)$ are closed linear subspaces of $B(X)$, and the operation L defined on $B_w(X)$ to X by setting $L(s)$ equal to the weak sum of s for each $s \in B_w(X)$ is linear and has norm* 1.

Proof. To show that $B_w(X)$ is closed in $B(X)$ suppose s_n is a sequence in $B_w(X)$ which converges to $s \in B(X)$. For each $n \in N$ let x_n denote the weak sum of s_n. Since $\{s_n\}$ is a Cauchy sequence in $B(X)$ there is for each $\epsilon > 0$ a natural number n_ϵ such that $\|s_n - s_m\| < \epsilon/2$ if $n, m > n_\epsilon$. For $n, m > n_\epsilon$ and $f \in X^*$ with $\|f\| < 1$ one has

$$|f(x_m - x_n)| < \textstyle\sum_i|f(s_n(i) - s_m(i))| < 2\|s_n - s_m\| < \epsilon,$$

the second inequality given by Lemma 2.1. It follows that $\{x_n\}$ is a Cauchy

sequence and therefore has a limit x. Again, suppose $\epsilon > 0$ is given and $f \in X^*$ with f non-zero. There is an n_ϵ such that

$$||s_n - s|| < \epsilon/(4||f||) \qquad n > n_\epsilon,$$

and since x_n converges to x, n_ϵ may be chosen large enough so

$$||x - x_n|| < \epsilon/(2||f||) \qquad n > n_\epsilon.$$

Hence, if $n > n_\epsilon$ then

$$|f(x) - \sum f(s(i))| < |f(x) - f(x_n)| + \sum_i |f(s_n(i) - s(i))|$$
$$< ||f||(\epsilon/(2||f||)) + 2||f||\,||s_n - s|| < \epsilon,$$

using Lemma 2.1 to get the second inequality. This proves that x is the weak sum of s.

To show that $B_s(X)$ is closed in $B(X)$ suppose $\{s_n\}$ is a sequence in $B_s(X)$ which converges to $s \in B(X)$. For each $n \in N$ let x_n denote the sum of s_n. Since $B_s(X) \subset B_w(X)$ and $B_w(X)$ is closed, s has a weak sum x. Also $\{x_n\}$ converges to x. Since $\{x_n\}$ converges to x and $\{s_n\}$ converges to s, if $\epsilon > 0$ is given there is $p \in N$, dependent on ϵ, such that $||x - x_p|| < \epsilon/3$ and $||s_p - s|| < \epsilon/3$. Also since $x_p = \sum_i s_p(i)$, there is a $q \in N$ such that if $r > q$ then

$$\left|\left|x_p - \sum_{i=1}^{r} s_p(i)\right|\right| < \epsilon/3.$$

Hence if $r > q$, then

$$\left|\left|x - \sum_{i=1}^{r} s(i)\right|\right| < ||x - x_p|| + \left|\left|x_p - \sum_{i=1}^{r} s_p(i)\right|\right|$$
$$+ \left|\left|\sum_{i=1}^{r} s_p(i) - \sum_{i=1}^{r} s(i)\right|\right| < \epsilon.$$

This shows that x is the sum of s.

It remains to show that L is a linear operation with norm 1. Let

$$E = [f: f \in X^* \text{ and } ||f|| = 1].$$

Fix $s \in B_w(X)$ and let $x = L(s)$. Then

$$||x|| = \sup[|f(x)|: f \in E] = \sup\left[\lim_{n\to\infty}\left|\sum_{i=1}^{n} f(s(i))\right| : f \in E\right]$$
$$< \sup\left[\sup\left\{\left|f\left(\sum_{i=1}^{n} s(i)\right)\right| : n \in N\right\} : f \in E\right]$$
$$= \sup\left[\sup\left\{\left|f\left(\sum_{i=1}^{n} s(i)\right)\right|\right\} : f \in E : n \in N\right]$$
$$= \sup\left[\left|\left|\sum_{i=1}^{n} s(i)\right|\right| : n \in N\right] < ||s||.$$

Hence L, which is obviously additive, is continuous and $||L|| < 1$. Since for any $x_0 \in X$ the sequence $\{x_0, \theta, \theta, \ldots, \theta, \ldots\}$ is in $B_w(X)$ and has x_0 for its norm, clearly $||L|| = 1$.

3. Extension of a theorem of Hadwiger to B-spaces. The following theorem is obtained by applying a modification of Hadwiger's argument **(5)** to the general case.

THEOREM 3.1. *If X is a B-space the following are equivalent:*

(i) *X has infinite dimension.*

(ii) *the difference $IS(X) \sim B(X)$ is non-void.*

(iii) *$U(X)$ is a proper subset of $IS(X)$.*

Proof. Because of the well-known fact that $U(X) \subset IS(X) \cap B(X)$ for all X, it is evident that (ii) implies (iii). Since $U(X) = IS(X)$ if X has finite dimension, (iii) implies (i). It will now be shown that (i) implies (ii). By a remark of Banach's **(1**, p. 238**)**, X contains a closed infinite dimensional linear subspace X_0 which has a basis $\{x(i)\}$ with $\|x(i)\| = 1$, $i \in N$. Using a result of Banach **(1**, pp. 110–111**)**, there is a sequence $\{f_i\}$ in X^* such that $f_i(x(j)) = \delta_{ij}$ and for each $x \in X_0$, $x = \sum_i f_i(x)x(i)$.

Consider the sequence of finite blocks

$$B_k = \{x(k)/k, -x(k)/k, \ldots, x(k)/k, -x(k)/k\}, \qquad k = 1, 2, 3, \ldots$$

where B_k consists of $2k^2$ terms each of which is either $x(k)/k$ or $-x(k)/k$ according as it is in an odd or an even place in B_k. Note that $x(k)/k$ occurs k^2 times in each B_k so the sum of the odd place terms in B_k has norm k. Construct a sequence s in X by adjoining the second block of terms to the first, the third block to this, etc. Since the norm of the sum of the odd place terms in each block is k, $s \notin B(X)$. Clearly $\sum_i s(i) = \theta$. It remains to show that s has an invariant sum. Suppose that s' is a rearrangement of s and that $y = \sum_i s'(i)$. Since X_0 is closed, $y \in X_0$. Express y by its biorthogonal development $y = \sum_i f_i(y)x(i)$. For arbitrary $i \in N$, we have $f_i(y) = \sum_j f_i(s'(j))$. Take n_0 large enough so that all terms in the block B_i occur in the sum

$$s'(1) + s'(2) + \ldots + s'(n_0).$$

If $n > n_0$ then

$$\sum_{j=1}^{n} f_i(s'(j)) = f_i\Big(\sum_{j \in F} s'(j)\Big) + \sum_{j \in F'} f_i(s'(j))$$

where $F = [j : j < n$ and $s'(j)$ is a term of $B_i]$ and

$$F' = [j : j < n \text{ and } j \notin F].$$

Now $\sum_F s'(j) = \theta$, and by biorthogonality $f_i(s'(j)) = 0$ if $j \in F'$, so $f_i(y) = 0$. Since $f_i(y) = 0$ for all i it follows that $y = \theta$.

4. Comparison of subspaces of $B(X)$**.** For any B-space X, $U(X) \subset B(X)$ so clearly $U(X) \subset B_s(X)$. Also $B_s(X) \subset IS(X)$ for any B-space X, because if $s \in B_s(x)$ and s has the sum x and if s' is a rearrangement of s with sum x' it follows that $f(x) = f(x')$ for all $f \in X^*$ so $x = x'$. With these observations the following lemma is obvious.

LEMMA 4.1. *For any B-space X,* $U(X) \subset B_s(X) = IS(X) \cap B(X) \subset B_w(X) \subset B(x)$.

A B-space X is *weakly complete* if and only if every weakly convergent sequence in X is weakly convergent to an element of X.

THEOREM 4.2. *If X is weakly complete then*

$$U(X) = B_s(X) = IS(X) \cap B(X) = B_w(X) = B(X) \subset IS(X).$$

The containment is proper if and only if X has infinite dimension.

Proof. For any B-space, $U(X) \subset IS(X)$ and it is well known that when X is weakly complete that $U(X) = B(X)$. Hence $B(X) \subset IS(X)$ when X is weakly complete. The theorem then follows by Lemma 4.1 and Theorem 3.1.

LEMMA 4.3. *If for a B-space X, U(X) is a proper subspace of B(X), then U(X) is a proper subspace*[1] *of* $B_s(X)$.

Proof. Suppose $s \in B(X) \sim U(X)$. For each $k \in N$ let B_k denote a block of $2k$ terms as follows:

$$B_k = \{s(k)/k, -s(k)k/, \ldots, s(k)/k, -s(k)/k\}.$$

that is, the even place terms in B_k are $s(k)/k$ and the odd place terms are $-s(k)/k$. We construct $s' \in B_s(X) \sim U(X)$ by adjoining the terms of the block B_2 to those of B_1 and then adjoining the terms of B_3 to these, etc. Clearly $\theta = \sum_i s'(i)$ and for each $f \in X^*$,

$$\sum_i |f(s'(i))| = 2\sum_i |f(s(i))| < \infty,$$

so $s' \in B_s(X)$. Finally, since $s \notin U(X)$ it follows that the series $\sum_i s'(i)$ has a subseries, namely, $\sum_i s'(2i - 1)$ which does not converge unconditionally. Hence $s' \notin U(X)$.

COROLLARY 4.4. *The B-space* $U(c_0)$ *is a proper subspace of* $B_s(c_0)$.

Proof. Consider the sequence $\{s_n\}$ in c_0 where for each n, $s_n(i) = 1$ if $i = n$ and $s_n(i) = 0$ if $i \neq n$. The sequence $\{s_n\}$ is an element of $B(c_0)$ but it does not have a sum so is not an element of $U(c_0)$. The corollary follows by Lemma 4.3.

LEMMA 4.5. *If for a B-space X, U(X) is a proper subspace of* $B_s(X)$ *then* $B_s(X)$ *is a proper subspace of* $B_w(X)$.

Proof. If $s \in B_s(X) \sim U(X)$ then there is a permutation t of N such that the sequence $\{s(t(i))\}$ does not have a sum. Let x denote the sum of s. Then x is the weak sum of s and since $s \in B(X)$ it follows that x is the weak sum of $\{s(t(i))\}$.

By Corollary 4.4 and Lemma 4.5 we have the next corollary.

COROLLARY 4.6. *The space* $B_s(c_0)$ *is a proper subspace of* $B_w(c_0)$.

LEMMA 4.7. *If for a B-space X, $U(X)$ is a proper subset of $B(X)$ then $B_w(X)$ is a proper subset of $B(X)$.*

Proof. By hypothesis there exists an $s \in B(X) \backsim U(X)$. Using a result of Orlicz (**1**, (3) on p. 270), there is a strictly increasing sequence t of natural numbers such that the sequence $\{s(t(i))\}$ does not have a weak sum. However it obviously inherits the property of belonging to $B(X)$ from s.

COROLLARY 4.8. *The space $B_w(c_0)$ is a proper subspace of $B(c_0)$.*

Proof. Since $B(c_0) \backsim U(c_0)$ is non-void the conclusion follows by Lemma 4.7.

Putting together the preceding corollaries we have the following

THEOREM 4.9. *For the B-space c_0, $U(c_0) \subset B_s(c_0) \subset B_w(c_0) \subset B(c_0)$, and* each containment is proper.

REFERENCES

1. S. Banach, *Théorie des opérations linéaires* (Warsaw, 1932)[1].
2. G. Birkhoff, *Integration of functions with values in a Banach space*, Trans. Amer. Math. Soc., *38* (1935), 357–378.
3. N. Dunford, *Uniformity in linear spaces*, Trans. Amer. Math. Soc., *44* (1938), 305–356.
4. I. Gelfand, *Abstrakte Funktionen und lineare Operatoren*, Mat. Sbornik, N.S., *46* (1938), 235–284.
5. H. Hadwiger, *Über die konvergenzarten unendlicher reihen in Hilbertschen raum*, Math. Zeit., *47* (1941), 325–329.
6. B. J. Pettis, *On integration in vector spaces*, Trans. Amer. Math. Soc., *44* (1938), 277–304.

[1]The author is indebted to the referee for the present form of Lemma 4.3 which is simpler and more general than the original.

Received March 25, 1955; in revised form October 28, 1955. This paper is from Chapter V of the author's dissertation, *On unconditional summability of sequences in semi-groups with a topology*, Tulane University, August, 1954. It was done under Contract N7-onr-434, Task Order III, Navy Department, Office of Naval Research. The author thanks Professor B. J. Pettis for helpful suggestions regarding this paper.

Alabama Polytechnic Institute

ON A QUASI-LINEAR EQUATION

RICHARD BELLMAN

1. Introduction. The purpose of this note is to establish some limit theorems for the non-linear recurrence relations

$$x_i(n+1) = \operatorname*{Max}_{q} \sum_{j=1}^{N} a_{ij}(q)\, x_j(n), \qquad i = 1, 2, \ldots, N; n > 0, \tag{1.1}$$

under certain assumptions concerning the initial values $c_i = x_i(0)$, and the coefficient matrices $A(q) = (a_{ij}(q))$.

Equations of this type occur in various parts of the theory of dynamic programming, as we shall indicate below, and are, in addition, of interest in furnishing a link between the theory of linear and non-linear operations, as we have discussed elsewhere **(1)**.

Generally speaking, these equations arise in the consideration of processes of Markoff type, see **(2)**, in which decisions are made at various stages of the process.

Results corresponding to those obtained below hold for the more general equations of the form

$$x_i(n+1) = \begin{cases} \operatorname*{Max}_{q} \sum_{j=1}^{N} a_{ij}(q)\, x_j(n), & i = 1, 2, \ldots, K < N, \\ \sum_{j=1}^{N} a_{ij}(q^*)\, x_j(n), & i = K+1, \ldots, N, \end{cases} \tag{1.2}$$

where q^* in the lower equations is determined by the upper equations.

2. The homogeneous equation. Let us consider the equation

$$\lambda y_i = \operatorname*{Max}_{q} \sum_{j=1}^{N} a_{ij}(q)\, y_j \qquad (i = 1, 2, \ldots, N), \tag{2.1}$$

where we impose the following conditions:

2.2 (a) $q = (q_1, q_2, \ldots, q_N)$ runs over some set of values, S, with the property that the maximum is attained in (1),

(b) $\infty > m > a_{ij}(q) > 0$ $(i, j = 1, 2, \ldots, N)$ for $q \in S$,

(c) for any q, let $\phi(q)$ denote the characteristic root of $A(q) = (a_{ij}(q))$ of largest absolute value, the Perron root, known to be positive. We assume that there exists at least one value of q for which $\phi(q)$ assumes its maximum for $q \in S$.

Received July 20, 1955.

We shall now prove

THEOREM 1. *Under these conditions, there exists a unique positive* λ *with the property that* 2.1 *has a positive solution,* $y_i > 0$ $(i = 1, 2, \ldots, N)$. *This solution is unique up to a multiplicative constant, and*

2.3 $$\lambda = \operatorname*{Max}_{q \epsilon S} \phi(q).$$

Proof. We begin by showing the existence of a positive λ and a positive set of solutions $\{y_i\}$. Consider the region defined by

$$y_i > 0, \ \sum_{i=1}^{N} y_i = 1.$$

The normalized transformation

2.4 $$y'_i = \left[\operatorname*{Max}_q \sum_{j=1}^{N} a_{ij}(q)\, y_j\right] \Big/ \left[\sum_{i=1}^{N} \operatorname*{Max}_q \sum_{j=1}^{N} a_{ij}(q)\, y_j\right],$$

is a continuous mapping of this region into itself. Hence there exists a fixed point, $\{y_i\}$. This fixed point is a solution of 2.1, with λ the denominator in 2.4. Each component y_i is positive because of the positivity of $a_{ij}(q)$.

To show that this solution is unique up to a multiplicative constant, let $[\mu, z]$ be another solution of 2.1 with $\mu > 0$ and z a positive vector. Let $\{q\}$ be the set of values for which the maximum is attained in 2.1 and $\{\bar{q}\}$ the similar set associated with z. Observe that we may have different sets for each i. We have then

2.5 $$\lambda y_i = \sum_j a_{ij}(q)\, y_j > \sum a_{ij}(\bar{q})\, y_j, \qquad i = 1, 2, \ldots, N,$$
$$\mu z_i = \sum_j a_{ij}(\bar{q})\, z_j.$$

Let us now assume, without loss of generality that $\lambda < \mu$. Let ϵ be a positive constant chosen so that one, at least, of the components $y_i - \epsilon z_i$ is zero, one at least is positive, and the others are non-negative. This can always be accomplished if y and z are not proportional. If i is an index for which $y_i - \epsilon z_i$ is zero, we have

2.6 $$0 = \mu(y_i - \epsilon z_i) > \lambda y_i - \epsilon \mu z_i > \sum_{j=1}^{N} a_{ij}(\bar{q})(y_j - \epsilon z_j) > 0,$$

since $a_{ij}(\bar{q}) > 0$, a contradiction. Hence $\lambda = \mu$, and y and z are proportional.

To show that $\lambda = \operatorname{Max} \phi(q)$, we proceed as follows. It is clear that λ, as the characteristic root of some $A(q)$, satisfies the inequality $\lambda < \mu$, where $\mu = \operatorname{Max} \phi(q)$. Assume that actually $\lambda < \mu$. Let $z = (z_1, z_2, \ldots, z_n)$ be a positive characteristic vector associated with μ and $\bar{q}$ a set of q-values which yield $\mu = \phi(\bar{q})$. Then we have

2.7 $$\mu z_i = \sum_{j=1}^{N} a_{ij}(\bar{q})\, z_j < \operatorname*{Max}_q \sum_{j=1}^{N} a_{ij}(q)\, z_j.$$

Since y_i is positive, we can find a positive constant m such that $z_i < my_i$ for $i = 1, 2, \ldots, N$. Hence 2.1 yields

$$\mu z_i < m \operatorname*{Max}_q \sum_{j=1}^{N} a_{ij}(q)\, y_j = m\lambda y_i. \tag{2.8}$$

Thus $z_i < my_i\,\lambda/\mu$. Iterating this, we obtain $z_i < my_i\,(\lambda/\mu)^k$, for arbitrary k. Since $\lambda/\mu < 1$, by assumption, this yields $z_i = 0$, a contradiction. Hence $\lambda = \mu$.

3. The recurrence relation. Let us now return to the recurrence relation of 1.1 and prove

THEOREM 2. *If, in addition to the conditions of 2.2, we assume that there is a unique q for which the maximum value of $\phi(q)$ is attained and that $c_i > 0$, then*

$$x_i(n) \sim ay_i\,\lambda^n, \tag{3.1}$$

as $n \to \infty$, where a is a constant dependent upon the initial values c_i.

Proof. Let us take $c_i > 0$ without loss of generality. There are then two positive constants k and K such that $ky_i < c_i < Ky_i$ $(i = 1, 2, \ldots, N)$. Let us show inductively that

$$ky_i\,\lambda^n < x_i(n) < Ky_i\,\lambda^n. \tag{3.2}$$

Assume that we have the result for n, then

$$\begin{aligned} x_i(n+1) &< K\lambda^n \operatorname*{Max}_q \sum_{j=1}^{N} a_{ij}(q)\, y_j = K\lambda^{n+1} y_i \\ &> k\lambda^n \operatorname*{Max}_q \sum_{j=1}^{N} a_{ij}(q)\, y_j = k\lambda^{n+1} y_i. \end{aligned} \tag{3.3}$$

To establish the asymptotic behavior we show that for n sufficiently large the set of q's which furnish the maximum in 1.1 is precisely the set which yields $\lambda = \operatorname{Max} \phi(q)$.

Assume the contrary. This means that infiitely often we employ a set $\{\bar{q}\}$ which is not identical with the q which furnishes the maximum in $\phi(q)$.

We then have, for $i = 1, 2, \ldots, N$,

$$x_i(n+1) = \sum_{j=1}^{N} a_{ij}(\bar{q})\, x_j(n) < \left(\sum_{j=1}^{N} a_{ij}(\bar{q})\, y_j \right) K\lambda^n. \tag{3.4}$$

For some index i we must have

$$\sum_{j=1}^{N} a_{ij}(\bar{q})\, y_j < \lambda y_i, \tag{3.5}$$

with strict inequality. For if

$$\sum_{j=1}^{N} a_{ij}(\bar{q})\, y_j > \lambda y_i$$

for all i, the characteristic root of $A(\bar{q}) = (a_{ij}(\bar{q}))$ of largest absolute value, $\phi(\bar{q})$, would at least equal $\lambda = \text{Max}\, \phi(q)$, which would contradict the assumption concerning the uniqueness of the maximum of $\phi(q)$.

Hence, for some component, say the first, we have

$$3.6 \qquad x_1(n+1) \leq \theta K\lambda^{n+1}y_1, \qquad 0 < \theta < 1.$$

Since $a_{1j}(q^*) > 0$ for i, j, where q^* is the value of q for which $\lambda = \phi(q^*)$, we see that, for $i = 1, 2, \ldots, N$,

$$3.7 \qquad x_i(n+2) \leq K\lambda^{n+1}\left[\sum_{j=2}^{N} a_{ij}(q^*)\, y_i + \theta a_{1j}(q^*)\, y_1\right] \leq \theta_1 K\lambda^{n+2}y_i,$$

where $\theta < 1$.

If therefore a set of q's distinct from q^* are used R times, we obtain

$$3.8 \qquad x_i(n) \leq \theta_1{}^R K\lambda^n y_i,$$

for n sufficiently large. Since $0 < \theta_1 < 1$, if R is too large we eventually contradict the lower bound for $x_i(n)$.

Hence for $n \geq n_0 = n_0(c_i)$, we have

$$3.9 \qquad x(n+1) = A(q^*)\, x(n),$$

whence the asymptotic statement of 3.1 follows.

4. A dynamic programming problem. Suppose that we are engaged in a multi-stage decision process of the following type. At each stage we have our choice of various operations, which we number $i = 1, 2, \ldots, K$. The ith operation has a probability distribution attached with the following properties:

4.11 There is a probability p_{ik} that we receive k units and the process continues, $k = 1, 2, \ldots, R$;

4.12 There is a probability p_{i0} that we receive nothing and the process terminates.

How do we proceed so as to maximize the probability that we receive at least n units before the process terminates?

Let us define the sequence

4.2 $u(n)$ = the probability of attaining at least n units before the termination of the process using an optimal procedure.

Then using the intuitive "principle of optimality" (1), we see that $u(n)$ satisfies the recurrence relation

$$4.3 \qquad u(n) = \begin{cases} \underset{i}{\text{Max}}\left[\sum_{k=1}^{R} p_{ik}u(n-k)\right], & n > 0, \\ 1, & n \leq 0. \end{cases}$$

Using methods similar to those above, we see that for large n,

4.4 $$u(n) \sim c\rho^n,$$

where ρ is the root of largest absolute value, necessarily positive, of

4.5 $$1 = \sum_{k=1}^{R} p_{ki}\rho^{-k},$$

for the value of i which maximizes ρ.

5. An analogue of a result of Markoff. Markoff showed that if

5.1 $$x_i(n+1) = \sum_{j=1}^{N} a_{ij}x_j(n) \qquad (n = 0, 1, \ldots)$$

and $x_i(0) > 0$, with the conditions

5.2 $$a_{ij} > 0, \ \sum_j a_{ij} = 1, \qquad (i = 1, 2, \ldots, N),$$

then

5.3 $$\lim_{n\to\infty} x_i(n) = c, \qquad (i = 1, 2, \ldots, N),$$

where c depends on the initial values.

The same proof shows that the same result holds for the sequence defined by

5.4 $$x_i(n+1) = \operatorname*{Max}_q \sum_{j=1}^{N} a_{ij}(q)\, x_j(n),$$

provided that the conditions in 5.2 hold uniformly in q. The constant will, of course, in general, be different from that above.

References

1. R. Bellman, *The theory of dynamic programming*, Bull. Amer. Math. Soc., *60* (1954), 503–516.
2. W. Feller, *An introduction to probability theory and its applications* (New York, 1950).

Rand Corporation,
Santa Monica, California

MODIFIED BOUNDARY VALUE PROBLEMS FOR A QUASI-LINEAR ELLIPTIC EQUATION

G. F. D. DUFF

1. Introduction. The quasi-linear elliptic partial differential equation to be studied here has the form

$$\Delta u = -F(P, u). \tag{1.1}$$

Here Δ is the Laplacian while $F(P, u)$ is a continuous function of a point P and the dependent variable u. We shall study the Dirichlet problem for (1.1) and will find that the usual formulation must be modified by the inclusion of a parameter in the data or the differential equation, together with a further numerical condition on the solution.

The negative sign on the right in (1.1) is included for convenience and also to emphasize that the behaviour of the right side will be the opposite of that usually studied. We shall generally take $F(P, u)$ to be a positive increasing function of u, these conditions being motivated by the following physical problem. Consider an equilibrium distribution of heat in a medium where the source density of heat generated depends on temperature u:

$$\rho = \rho(u) = F(P, u).$$

That ρ and hence $F(P, u)$ in (1.1) should be positive and increasing with u is a natural assumption.

The known results for quasi-linear equations such as (1.1) are, roughly speaking, of two kinds: local theorems, and in-the-large existence proofs for equations

$$\Delta u = +F(P, u) \tag{1.2}$$

where $F(P, u)$ is an increasing function of u. By local theorems are meant those in which some restriction of size is placed on the boundary values, the domain, or the non-linearity of the function $F(P, u)$. Among these we might include the case when $F(P, u)$ is bounded independently of u. The Dirichlet theorem and various other boundary value results have been proved in such circumstances. (**2**; **4**, vol. II, Ch. V; **6**, Ch. II).

On the other hand, global existence theorems for (1.2) have been found by many authors. (**3**; **6**, Ch. II). The possibility of this may be recognized if one

Received August 11, 1955. This paper was written at the 1955 Summer Research Institute of the Canadian Mathematical Congress. The author's thanks are due to the National Research Council of Canada for a fellowship held at this time. He is also indebted to Professor J. Leray for a stimulating discussion and helpful comments.

constructs the equation of variation of (1.2) with respect to an external parameter: this variational equation has the linear form

$$(1.3) \qquad \Delta v = F_u(P, u)\, v,$$

with positive coefficient $F_u(P, u)$. Such equations satisfy a maximum principle in the sense that the maximum absolute value of any solution is taken on the boundary. Thus a priori estimates can be found for the solutions of (1.3) and hence for those of (1.2).

These methods will not apply to (1.1). Even in the linear case, it is evident that the usual statement of the Dirichlet problem, namely the assertion that a solution having given boundary values exist, does not hold unless $F(P, u)$ is restricted in some way. Indeed, if λ is an eigenvalue, solutions of $\Delta u + \lambda u = 0$ have boundary values restricted by one or more conditions of orthogonality. This particular case will be relevant to Theorem III below; we shall later furnish a similar example which pertains to the main Theorem I and which shows that the conventional Dirichlet problem is not then always solvable.

This discussion suggests that we should frame boundary value problems for (1.1) in such a way that some *a priori* bound can be included in the statement of the problem. We will show that in a certain sense it is sufficient to bound the solution from above. In fact we assume that the actual maximum of the solution has a stated value. If, however, one additional numerical condition is assigned, it is evident that a corresponding degree of freedom should be allowed for the boundary values of the solution. This we shall permit by introducing a parameter t, of the nature of an eigenvalue parameter, into the boundary condition. Thus the main theorem asserts the existence of a solution with a stated maximum and with boundary values proportional to a given function.

We then establish some variations of this theorem, allowing the parameter to appear in various ways in the differential equation instead of the boundary condition. These solutions have an assigned maximum together with given boundary values. We conclude with a Neumann boundary value theorem for an equation similar to (1.1) but containing an additional linear term.

2. Preliminaries. Let V_N be a Riemannian manifold of dimension N with positive definite metric of class C^4 in a given coordinate network:

$$ds^2 = a_{ik}\, dx^i dx^k;$$

then the Laplace operator has the form

$$(2.1) \qquad \Delta u = \frac{1}{\sqrt{a}} \frac{\partial}{\partial x^i}\left(\sqrt{a}\, a^{ik} \frac{\partial u}{\partial x^k}\right).$$

where $a = |a_{ik}|$ and the associate tensor a^{ik} satisfies

$$a^{ik} a_{kj} = \delta^i{}_j.$$

We consider a compact domain D of V_N, having a boundary surface B of class C^2 in the above coordinate system. Points of D will be denoted by capitals

$P, Q, \ldots$ and points of B by lower case $p, q, \ldots$. We assume that $F(P, u)$ is a continuous function of P and u together; other conditions appropriate to each theorem will be stated separately. All functions and parameters used are real-valued.

The existence proofs which follow will be based on the Schauder-Leray theorem **(5)**, which we will state here. We work with the separable Banach space C, of continuous functions on the closure of the domain D, with the norm

$$||u|| = \max_{P \epsilon D} |u(P)| . \tag{2.2}$$

Let Ω be a bounded domain of C, with boundary Ω', and let $T_k[u]$ be an operator defined in $\Omega + \Omega'$ which satisfies the conditions

(a) $T_k[u]$ is jointly uniformly continuous in k and u, for $0 < k < 1$ and $u \in \bar{\Omega} = \Omega + \Omega'$.

(b) $T_k[u]$ is a compact or completely continuous operator, transforming bounded sets into compact sets **(1**, Ch. VI**)**. Suppose also that the equation

$$u = T_k[u] \tag{2.3}$$

has no solution on Ω' for $0 < k < 1$, and that for $k = 0$ it has a solution in Ω. Finally, let

$$v = u - T_0[u]$$

be a homeomorphism of C. Then the conclusion of the Schauder-Leray theorem is that *the equation*

$$u = T_1[u] \tag{2.4}$$

has at least one solution in Ω.

Separate choices for $T_k[u]$ and for Ω will be made in each of the following theorems. In each case condition (b) above is satisfied essentially because the integral operator with kernel the Green's function of D for $\Delta u = 0$ is completely continuous in the space C. We include a demonstration of this in the proof of Theorem I.

3. The modified Dirichlet problem. Let M be a positive number given in advance, and let $f(p)$ be a C^1 function on the boundary B which is positive:

$$0 < m_0 < f(p) < M_0 < \infty . \tag{3.1}$$

We also take $m_0 < 1$, which is always possible, for a reason which will appear later. Let $F(P, u)$ be a positive non-decreasing function of u.

THEOREM I. *There exists a solution of* (1.1) *with maximum value M and boundary values proportional to $f(p)$.*

The constant of proportionality being denoted by t, we have

$$u(p) = tf(p) \tag{3.2}$$

and also

$$\max_{P \epsilon D} u(P) = M \tag{3.3}$$

To establish the existence of such a solution, we begin by constructing the harmonic function $v_0(P)$ with boundary values $f(p)$. Thus

$$\Delta v_0(P) = 0, \quad v_0(p) = f(p), \tag{3.4}$$

and in view of the maximum principle for harmonic functions and (3.1) we have

$$m_0 < \min f(p) < v_0(P) < \max f(p) = M_0, \tag{3.5}$$

these inequalities holding for $P \in D + B$.

Now a solution of (1.1) with boundary values $tf(p)$ satisfies the integral equation

$$u(P) = \int_D G(P, Q)\, F(Q, u(Q))\, dV_Q + tv_0(P), \tag{3.6}$$

where $G(P, Q)$ is the harmonic Green's function of the domain D. We note that $G(P, Q)$ is non-negative **(2)**. Conversely, a solution of (3.6) is actually a solution of (1.1) with boundary values $tf(p)$, as may be verified by operating on (3.6) with the Laplacian, and noting that the integral on the right has vanishing boundary values. We observe that to satisfy the maximum condition (3.3) we must make an appropriate choice of t, which will in turn depend on $u(P)$, so that a fixed value for t cannot be determined at this stage.

We therefore define the non-linear functional

$$T_k[u](P) = \int_D G(P, Q)\, F(Q, ku(Q))\, dV_Q + t_k[u]\, v_0(P), \tag{3.7}$$

where $t = t_k[u]$ is so chosen that

$$\max_{P \epsilon D+B} T_k[u](P) = M. \tag{3.8}$$

Since $v_0(P)$ satisfies (3.5) we see that such a choice of t is always possible, since the right side of (3.7) is a strictly increasing function of t tending to $\pm\infty$ with t.

We now show that $t_k[u]$ is bounded, provided that $0 < k < 1$ and $u < K$, where K is a fixed constant. Let

$$A = \max_{P \epsilon D} \int_D G(P, Q)\, F(Q, K)\, dV_Q; \tag{3.9}$$

this number exists and is positive. Now if $A < M$ it would appear that $t_k[u]$ in (3.7) should be positive. However it is clear that

$$t_k[u] < M/M_0,$$

since $M_0 = \max v_0(P)$ and $T_k[u] < M$. This furnishes an upper bound for $t_k[u]$. If $M < A$, $t_k[u]$ may be negative; however

$$\int_D G(P, Q)\, F(Q, ku(Q))\, dV_Q < A,$$

since $ku < K$ for $0 < k < 1$ and F is a non-decreasing function of u. Thus the multiple of $v_0(P)$ required to reduce the maximum of $T_k[u]$ to M does not exceed $(A - M)/m_0$. We therefore have

$$-\frac{A - M}{m_0} < t_k[u] < \frac{M}{m_0}. \tag{3.10}$$

This shows that if u is bounded above, $t_k[u]$ is bounded below and in fact bounded. The lower bound depends on A and hence on the upper bound K of u.

Since the integral in (3.7) is non-negative, it follows that $T_k[u]$ is bounded below:

$$-\frac{A - M}{m_0} M_0 < T_k[u]. \tag{3.11}$$

Combining (3.8) and (3.11), we see that $T_k[u]$ is bounded in both directions.

To apply the Schauder-Leray theorem, we set

$$K = 2M/m_0 > M,$$

and let A be defined by (3.9) with this value of K. Then we choose Ω to be the connected domain of C defined by

$$\Omega: -\frac{2M}{m_0}|A - M| - \epsilon_0 < u(P) < \frac{2M}{m_0}, \qquad \epsilon_0 > 0. \tag{3.12}$$

The boundary Ω' consists of those functions u for which equality holds on either side for one or more points of $D + B$. Now $T_k[u]$ is defined on $\Omega + \Omega'$ and is continuous in both k and u. This is easily verified since $F(P, ku)$ is uniformly continuous in ku, and the integral

$$\int_D G(P, Q)\, dV_Q$$

is a continuous function of P, vanishing on B, and so is bounded. Thus the integral in (3.7) depends continuously on ku and so, therefore, does $t_k[u]$. Hence $T_k[u]$ is continuous in k and u together.

We now show that $T_k[u]$ is a compact operator in C. Let $\{u_n\}$ be a uniformly bounded sequence of continuous functions. From (3.8) and (3.11) we see that $T_k[u_n]$ is bounded, independently of n and P. We now show that the sequence $T_k[u_n](P)$ is equicontinuous in P by forming the difference

$$\begin{aligned} |T_k[u_n](P_2) - T_k[u_n](P_1)| &< +\, t_k[u_n]|\, v_0(P_2) - v_0(P_1)| \\ &+ \int_D |G(P_2, Q) - G(P_1, Q)|\, F(Q, u_n(Q))\, dV_Q \end{aligned} \tag{3.13}$$

Since u_n is bounded independently of n, so is $F(Q, u_n)$ and also $t_k[u_n]$: let F_0

and t_0 be bounds for the absolute values of these sequences. Thus the preceding difference is less than

$$F_0 \int_D |G(P_2, Q) - G(P_1, Q)| \, dV_Q + t_0 | v_0(P_2) - v_0(P_1)|.$$

The second term here tends to zero as $P_2 \to P_1$, since $v_0(P)$ is continuous. To estimate the integral containing the Green's functions, we suppose that the distance $s(P, Q) < \delta$ and denote by S_η a geodesic sphere of radius η about P_1. For $P \neq Q$, $G(P, Q)$ is continuous, and we can therefore choose δ so small that for $Q \in D - S_\eta$, the difference

$$|G(P_2, Q) - G(P_1, Q)| < \epsilon_1.$$

We then write

$$\begin{aligned} &\int_D |G(P_2, Q) - G(P_1, Q)| \, dV_Q \\ &< \int_{D - S_\eta} |G(P_2, Q) - G(P_1, Q)| \, dV_Q \\ &\quad + \int_{S_\eta} \{G(P_1, Q) + G(P_2, Q)\} \, dV_Q \\ &< \epsilon_1 \int_D dV_Q + \int_{\bar{S}} G(P_2, Q) \, dV_Q + \int_{S_\eta} G(P_1, Q) \, dV_Q. \end{aligned} \tag{3.14}$$

Here $\bar{S}$ is a sphere of radius 2η about P_1, which certainly contains S_η if $S(P_1P_2) < \eta$. Since

$$G(P, Q) \sim \frac{1}{\omega_N(N-2)} \, s(P, Q)^{-N+2}, \qquad P \to Q,$$

the integrals over small spheres converge like

$$\int_{S_\eta} G(P, Q) \, dV_Q \sim \frac{1}{2(N-2)} \, \eta^2 \to 0, \qquad \eta \to 0,$$

uniformly with respect to P in D. Given $\epsilon > 0$, we choose η so small that the second and third terms on the right in (3.14) are each less than $\frac{1}{4}\epsilon$. We can then choose $\delta < \eta$ so small that the first term is less than $\frac{1}{4}\epsilon$. Also for $s(P_2, P_1)$ sufficiently small the second term on the right of (3.13) can be made less than $\frac{1}{4}\epsilon$. This shows, finally, that the sequence $T_k[u_n](P)$ is equicontinuous, uniformly for P in $D + B$. By Ascoli's theorem **(1)**, the sequence contains a uniformly convergent subsequence with a continuous limit. That is, $T_k[u]$ is a compact operator in C.

Next we demonstrate that for $0 < k < 1$, the equation

$$u(P) = T_k[u](P) \tag{3.15}$$

has no solution lying on the boundary Ω'. Since for any solution,

$$\max u = \max T_k[u] = M,$$

we see from (3.12) and the condition $m_0 < 1$ that

$$u(P) < M < 2M/m_0 = K.$$

Since A was defined by (3.9) for this K, we see that if $t_k[u] < 0$, then

$$\begin{aligned} -\epsilon_0 - \frac{2M}{m_0}|A - M| < \frac{-M}{m_0}|A - M| &< M_0\, t_k[u] \\ &< t_k[u]\, v_0(P) \\ &< \int_D G(P, Q)\, F(Q, ku(Q))\, dV_Q + t_k[u]\, v_0(P) \\ &= T_k[u](P) = u(P). \end{aligned}$$

Hence the strict inequality on the left holds in (3.12) for any solution and so if $t_k[u] < 0$ no solution can lie in Ω'. If $t_k[u] > 0$, then $T_k[u] > 0$ and the same conclusion follows at once.

Now for $k = 0$ the equation (3.15) has a unique solution since the operator $T_0[u]$ is then independent of u. (Thus the mapping $v = u - T_0[u]$ is a homeomorphism). In fact the solution u for $k = 0$ is the solution of $\Delta u = -F(P, 0)$, with $\max u = M$ and $u(p) = t f(p)$.

From the Schauder-Leray theorem we may now conclude that (3.15) has a solution for each k, $0 < k < 1$. For $k = 1$, we observe that in view of (3.7), (3.15) becomes equivalent to the integral equation (3.6). Thus the solution $u(P)$ for $k = 1$ satisfies (1.1) and has boundary values $t f(p)$. From (3.8) and (3.15) it follows that its maximum value is M. This completes the proof of the theorem.

Two minor extensions of this result will be noted here. First, we can treat the case where $F(P, u)$ is only bounded below:

$$F(P, u) > -K_1$$

by taking as a new variable $\bar{u} = u + v$, with v the solution of $\Delta v = -K_1$ which vanishes on B. Second, we may replace the boundary values $t f(p)$ by a more general continuous function $f(p, t)$ which is strictly increasing with t and tends to $\pm\infty$ with t.

4. Qualitative behaviour of the boundary values. The theorem of the preceding section would be of comparatively small interest if it were possible to solve the conventional Dirichlet problem which concerns the existence of a solution with given boundary values. We show that this problem is not solvable for the class of non-linear equations here considered.

Let λ_1 be the lowest Dirichlet eigenvalue of D for $\Delta u + \lambda u = 0$, and let the corresponding eigenfunction be denoted by u_1. From (**4**, vol. I, ch. VI, §6) we see that u_1 is of one sign in D, say non-negative. Hence the outward normal derivative $\partial u_1/\partial n$ is non-positive, and also does not vanish identically.

Now let u be any solution of (1.1) with boundary values $t\, f(p)$, and let us suppose that

$$F(P, u) > \lambda_1 u \tag{4.1}$$

for all values of u. *Then the value of t is necessarily negative.*

This assertion follows readily from Green's formula, since

$$\begin{aligned} t \int_B f\, u_{1n} dS &= \int_B (u\, u_{1n} - u_n u_1)\, dS \\ &= \int_D (u\Delta u_1 - u_1 \Delta u)\, dV \\ &= \int_D u_1[F(P, u) - \lambda_1 u]\, dV. \end{aligned} \tag{4.2}$$

Since $u_1 > 0$ in D the integral on the right is positive and since $\int_B u_{1n}\, dS < 0$ we conclude that $t < 0$. If in (4.1) the equality sign is permitted we would find $t \leqslant 0$; the case

$$F(P, u) = \tfrac{1}{2}\lambda_1(u + |u|) \tag{4.3}$$

illustrates this possibility.

Thus, if (4.1) holds, (1.1) can not have any solutions with positive boundary values. This shows that the conventional Dirichlet problem for (1.1) is impossible. Since in the physical interpretation of heat generation one would expect $F(P, u)$ to be a rapidly increasing function of u as $u \to +\infty$, it seems worthwhile to find the closest analogue of the conventional Dirichlet theorem for such equations. Though Theorem I is not the only variant which might be considered, it has physical meaning since:

(a) the maximum temperature is prescribed.

(b) the distribution (or ratio) of temperatures on the boundary is prescribed, so that if the actual boundary value is known at one point, all other boundary values are determined.

We continue the qualitative discussion of the values of t. If (4.1) holds only for

$$u > u_0 > 0, \tag{4.4}$$

we have

$$t < \frac{u_0}{m_0}, \quad m_0 = \min_{p \in B} f(p), \tag{4.5}$$

since otherwise we should have $t\, f(p) > u_0$ and, the minimum value of a solution of (1.1) being assumed on the boundary, this would lead to

$$u > u_0 \qquad \text{in } D.$$

But then (4.1) and (4.2) show that $t < 0$, which is a contradiction.

If we regard t as a function of M for fixed $f(p)$, we can show that t is a

continuous function of M. This follows from the Schauder-Leray theorem if we consider the functional

$$T_1[u](P) \equiv T_{1,M}[u](P)$$

in its dependence on M. We need only choose the domain Ω so that M is free to vary in a small interval and so that no solution of $u(P) = T_{1,M}[u](P)$ can cross the boundary of Ω. The reader will readily be able to supply the details here.

We now show that if

$$F_u(P, u) < \lambda_1, \qquad u < \bar{M}, \tag{4.6}$$

then t is a monotone strictly increasing function of M for $M < \bar{M}$. This will be established by finding a contradiction to the contrary assumption, which is that there exist M_1 and M_2, $M_1 < M_2 < \bar{M}$, such that $t_1 > t_2$. Let u_1 and u_2 be the respective solutions. Then $w = u_2 - u_1$ satisfies

$$\begin{aligned} \Delta w &= - F(P, u_2) + F(P, u_1) \\ &= - (u_2 - u_1)\, F_u(P, u_1 + \theta(P)(u_2 - u_1)) \\ &= - w\, F_u(P, u_3), \end{aligned}$$

say. Here u_3 is intermediate in value to u_1 and u_2, so $u_3 < \bar{M}$. Since

$$w = u_2 - u_1 = (t_2 - t_1) f(p) < 0 \qquad \text{on } B$$

and $w > M_2 - M > 0$ at the maximum of u_2, there exists a domain $D_1 \subseteq D$ wherein w is positive, and such that $w = 0$ on the boundary B_1 of D_1. Let λ' be the lowest Dirichlet eigenvalue of D_1; then (**4**, vol. I; ch. VI, §6) we have $\lambda_1 < \lambda'$ since $D_1 \subseteq D$. Let u_1' be the corresponding eigenfunction; we see as in (4.2) that

$$0 = \int_{D_1} u_1' w\, [F_u(P, u_3) - \lambda']\, dV,$$

and this is a contradiction since $u_1' > 0$, $w > 0$ and $F_u(P_3, u_3) < \lambda_1 < \lambda'$ in D_1, no one of the three factors vanishing in any open subset of D_1. This proves the results stated.

For example, if

$$F(P, u) = \begin{cases} 0, & u < 0, \\ u^n, & u > 0, \end{cases} \qquad n > 1,$$

we see that (4.1) holds for

$$u > u_0 = \lambda_1^{1/(n-1)}$$

and so an upper bound for t is known. For $M = 0$ the solution $u \equiv 0$ fulfills the conditions of Theorem I with $t = 0$. Since (4.6) holds for

$$u < \left(\frac{\lambda_1}{n}\right)^{1/(n-1)},$$

we see that t increases and is positive for

$$0 < M < \left(\frac{\lambda_1}{n}\right)^{1/(n-1)}.$$

The behaviour of t as $M \to \infty$ seems difficult to determine.

5. Related eigenvalue problems. The theorems of this section differ from the preceding result in that the parameter t appears in the differential equation instead of the boundary condition. They have therefore the character of eigenvalue problems, although the conditions to be fulfilled by the solution include the assigning of boundary values.

Let $F(P, u)$ be a continuous positive function, bounded away from zero:

$$F(P, u) \geqslant \delta > 0, \tag{5.1}$$

and consider the problem of finding a solution of

$$\Delta u = -tF(P, u) \tag{5.2}$$

with given boundary values $f(p)$ and a given maximum M. Let us assume that $f(p)$ is C^1 with maximum

$$M_0 = \max_B f(p). \tag{5.3}$$

Then without loss of generality we may take

$$M > M_0, \tag{5.4}$$

since in any case $M \geqslant M_0$ is necessary, while if $M = M_0$, we may take $t = 0$ (4.2) and find a harmonic solution of the problem.

Since a solution of the problem satisfies the integral equation

$$u(P) = t\int_D G(P, Q)\, F(Q, u(Q))\, dV_Q + v_0(P), \tag{5.5}$$

where $v_0(P)$ is again harmonic with boundary values $f(p)$, we define the new operator

$$T_k^1[u](P) = t_k^1[u]\int_D G(P, Q)\, F(Q, ku(Q))\, dV_Q + v_0(P), \tag{5.6}$$

with the choice of t governed by the condition

$$\max_{P\epsilon D} T_k^1[u](P) = M. \tag{5.7}$$

To show that this choice is possible we note that the non-negative integral

$$\int_D G(P, Q)\, dV_Q \tag{5.8}$$

has a maximum G_0 say for $P = P_0$ in D. Now for $t = 0$ the right side of (5.6) is less than M; consequently $t_k^1[u]$ must be positive. As t increases, so does the expression on the right in (5.6). However at $P = P_0$ we have

$$t\delta G_0 \leqslant t\delta\int_D G(P_0, Q)\, dV_Q$$

$$< t \int_D G(P_0, Q)\, F(Q, ku(Q))\, dV_Q$$
$$< M - v(P_0).$$

Let us denote by m_0 the minimum of $f(p)$, then by the maximum principle for harmonic functions

$$m_0 < v_0(P), \qquad P \in D,$$

and so we find

$$0 < t_k{}^1[u] < (M - |m_0|)\, \delta^{-1} G_0{}^{-1}. \tag{5.9}$$

Since $t_k{}^1[u]$ is positive, we have

$$m_0 < v_0(P) < T_k{}^1[u]$$

and therefore $T_k{}^1[u]$ has the bounds

$$m_0 < T_k{}^1[u] < M. \tag{5.10}$$

We now choose for Ω the connected region of C:

$$\Omega\colon m_0 - \epsilon < u < M + \epsilon, \tag{5.11}$$

and consider the equation

$$u = T_k{}^1[u], \tag{5.12}$$

for $0 < k < 1$. That $T_k{}^1[u]$ is jointly continuous in k and u is evident on inspection. To show that this operator is compact, we select from any bounded set of functions a subsequence $\{u_n\}$ such that $t_k{}^1[u_n]$ converges to a limit. This is possible on account of (5.9). A proof similar to that in the preceding sections shows that

$$\int_D G(P, Q)\, F(Q, u(Q))\, dV_Q$$

is compact, and the result follows if we consider the subsequence $\{u_n\}$.

For $0 < k < 1$, we see from (5.10) that (5.12) has no solutions on Ω', since this would contradict (5.11). For $k = 0$, $T_k{}^1[u](P)$ is independent of u and so (5.12) has a unique solution. The Schauder-Leray theorem now shows that for $k = 1$, (5.12) has a solution. Thus the integral equation (5.5) has a solution $u(P)$ with maximum M, and this establishes the result, which we state as follows.

THEOREM II. *There exists a solution for suitable t of*

$$\Delta u = -tF(P, u), \qquad F \geq \delta > 0,$$

with assigned boundary values $f(p) < M$ and maximum M.

The proof shows that the minimum value of the solution is attained on the boundary, and so is equal to m; however this could be deduced from the differential equation given that t is positive.

From our next theorem we insert the parameter t with the dependent variable

u in $F(P, u)$. This requires a different set of conditions to be satisfied by $F(P, u)$, namely

$$F(P, 0) = 0 \tag{5.13}$$

and

$$F_u(P, u) > \delta > 0. \tag{5.14}$$

Thus we consider the differential equation

$$\Delta u = - F(P, tu), \tag{5.15}$$

and look for a solution with maximum M and boundary values $f(p)$ where

$$0 < m_0 < f(p) < M_0 < M. \tag{5.16}$$

The necessity of these restrictions will appear; meanwhile we remark that the case $M_0 = M$ can be solved for $t = 0$ with a harmonic solution.

The appropriate integral equation is now

$$u(P) = \int_D G(P, Q)\, F(Q, tu(Q))\, dV_Q + v_0(P). \tag{5.17}$$

We shall supply the parameter k in front of the integral, but this leads to a minor difficulty which suggests the addition of a further term. We define

$$T_k^2[u](P) = k \int_D G(P, Q)\, F(Q, tu(Q))\, dV_Q + C(1 - k)\, t + v_0(P), \tag{5.18}$$

where

$$2C = \delta m_0 G_0,$$

and G_0 is again the maximum value of the integral (5.8). For $0 < k < 1$ the right side of (5.18) is an increasing function of t, and we can choose $t = t_k^2[u]$ so that

$$\max T_k^2[u] = M. \tag{5.19}$$

Since the first two terms in T_k^2 have the sign of t, and since $v_0(P) < M$, it is evident that $t_k^2[u]$ must be positive. Thus for $0 < k < 1$, $T_k^2[u]$ will have the lower bound m_0, since $m_0 < v_0(P)$. We therefore define the region Ω of function space C as

$$\Omega\colon 0 < \tfrac{1}{2}m_0 < u(P) < K, \tag{5.20}$$

where K is a large positive constant as yet not fixed, but which exceeds M.

To show that T_k^2 is completely continuous in $\Omega + \Omega'$ we need a uniform bound for $t_k^2[u]$, $u \in \Omega + \Omega'$. To find this, we take the point P_0 where (5.8) has maximal value $G_0 > 0$, and note that for $u \in \Omega$, $F(P, u) > \frac{1}{2}\delta m_0$. Then

$$\begin{aligned} M > T_k^2[u] &> \tfrac{1}{2}kt\delta m_0 G_0 + C(1 - k)\, t + m_0 \\ &= \tfrac{1}{2}\delta m_0 G_0 t + m_0 \end{aligned}$$

according to the definition of C in (4.18). Thus for $u \in \Omega$, we have

$$0 < t_k^2[u] < 2\,\frac{M - m_0}{\delta m_0 G_0}. \tag{5.21}$$

The conclusion now follows quickly from the Leray-Schauder theorem. The equation

$$u(P) = T_k^2[u](P) \tag{5.22}$$

has no solutions on Ω' for $0 < k < 1$, since

$$\tfrac{1}{2}m_0 < m_0 < T_k^2[u] < M < K.$$

For $k = 0$, the operator T_k^2 is independent of u, so that a unique solution exists. Thus for $k = 1$ the conclusion follows that (5.22) has a solution. From (5.18) we see that (5.17) is then satisfied.

THEOREM III. *Let $F(P, u)$ satisfy* (5.13) *and* (5.14). *Then there exists for a suitable value of t a solution of*

$$\Delta u = -F(P, tu)$$

with assigned maximum $M > 0$ in $D + B$ and given boundary values $f(p) < M$ on B.

We note that $F(P, u) = \lambda_1 u$, where λ_1 is the lowest eigenvalue as in §4, yields a counterexample to the solvability of the conventional Dirichlet problem for this equation, since an orthogonality condition is necessary.

We conclude this section with a similar theorem for the equation

$$\Delta u = -F(P, u) - t\rho(P). \tag{5.23}$$

Again the solution is to have a given maximum M and boundary values $f(p) < M$. The detailed assumptions are as follows. We take for $F(P, u)$ the restrictions

$$F(P, u) > -F_0 \tag{5.24}$$

and

$$F_u(P, u) > 0, \tag{5.25}$$

while the coefficient of t on the right in (4.23) must satisfy

$$\rho(P) > \rho_0 > 0. \tag{5.26}$$

The integral equation of the problem is

$$u(P) = \int_D G(P, Q)\,[F(Q, u(Q)) + t\rho(Q)]\,dV_Q + v_0(P), \tag{5.27}$$

and so, defining

$$R(P) = \int_D G(P, Q)\,\rho(Q)\,dV_Q > 0, \tag{5.28}$$

we set

$$T_k^2[u](P) = k\int_D G(P, Q)\,F(Q, u(Q))\,dV_Q + tR(P) + v_0(P). \tag{5.29}$$

The choice of $t = t_k^3[u]$ is again governed by

$$\max T_k^3[u] = M. \tag{5.30}$$

For the domain Ω we take

$$\Omega: -K < u < M + \epsilon, \tag{5.31}$$

where K is a large positive constant. Now for $u \in \Omega$ we have from (5.24) and (5.25) a limitation for $F(P, u)$:

$$|F(P, u)| < A. \tag{5.32}$$

Since $F(P, u)$ is bounded as $u \to -\infty$, A is independent of K.

We now obtain bounds for $t = t_k^3[u]$. Since $v_0(P) < M$, the first two terms together in (5.29) must be somewhere positive. Since $G(P, Q)$ is a non-negative kernel, this implies that the integrand

$$kF(Q, u(Q)) + t\rho(Q)$$

is somewhere positive. Hence at some point Q_1,

$$t\rho(Q_1) > -kF(Q_1, u(Q)) > -kA$$

and so

$$t > -kA/\rho_0.$$

This furnishes a lower bound for t. An upper bound may be found if we note that at the point P, where $R(P_1) = R_1$ is maximal, we have

$$tR_1 < M - k\int_D GF\,dV - v_0$$
$$< M + kG_0F_0 - m_0.$$

Thus

$$-A/\rho_0 < t < (M + G_0F_0 - m_0)/R_1, \tag{5.33}$$

and these bounds are independent of K.

The necessary lower bound for $T_k^3[u]$ is obtained by taking lower bounds for each term. Thus

$$T_k^3[u] > -F_0G_0 - \frac{A}{\rho_0}R_1 + m_0, \tag{5.34}$$

where m_0 is a lower bound for $f(p)$. This lower bound (4.34) is independent of K and so if we choose

$$K = 2\left(F_0G_0 + \frac{AR_1}{\rho_0} + |m_0|\right),$$

then the equation

$$u = T_k^3[u] \tag{5.35}$$

will have no solutions on Ω' for $0 < k < 1$. For $k = 0$ there is a unique solution

as before. For $k = 1$, there must accordingly exist a solution and from (5.29) we see that (5.27), is satisfied for a certain value of t. The maximum condition (5.30) also holds and the solution of the problem is thus completed.

THEOREM IV. *Let $F(P, u)$ satisfy (5.24) and (5.25), and let $\rho(P)$ satisfy (5.26). Then there exists for a suitable value of t a solution of*

$$\Delta u = - F(P, u) - t\rho(P),$$

with assigned maximum M in $D + B$ and given boundary values $f(P) < M$ on B.

The various conditions imposed on $F(P, u)$ in these theorems can be slightly relaxed in various ways. However it is to be noted that the conditions of Theorem III exclude all functions $F(P, u)$ satisfying the restrictions of the other theorems.

6. A modified Neumann problem. As an illustration of the way in which this method of proving existence theorems can be applied to other types of boundary condition, we include here a modified Neumann problem for the equation

$$\Delta u - \delta u = - F(P, u), \qquad \delta > 0, \tag{6.1}$$

where

$$F(P, u) > - F_0, \quad F_u(P, u) > 0. \tag{6.2}$$

The boundary condition shall be

$$\frac{\partial u}{\partial u} = g_0(p) + tg_1(p), \tag{6.3}$$

for some value of t. We take $g_0(p)$ and $g_1(p)$ to be C^1 with

$$g_1(p) > 0. \tag{6.4}$$

The usual maximum condition $\max u = M$ shall hold.

The Neumann function $N(P, Q)$ of the linear equation

$$\Delta u - \delta u = 0 \tag{6.5}$$

may be written as

$$N(P, Q) = G(P, Q) + K(P, Q), \tag{6.6}$$

where $G(P, Q)$ is the Green's function, and $K(P, Q)$ the Bergman kernel function, of (6.6). (2) We shall need the complete continuity in the space C of the operator with kernel $N(P, Q)$; this will be established by showing that the operators based on $G(P, Q)$ and $K(P, Q)$ are completely continuous. Indeed the proof for $G(P, Q)$ is the same as in §3. Now let us write down Green's first formula on D with argument functions $K(P, Q)$ and 1. Since $K(P, Q)$ is a solution of the differential equation, we get

$$\int_D [\nabla K \cdot \nabla 1 + \delta K \cdot 1]\, dV = \int_B 1 \frac{\partial K}{\partial n}\, dS.$$

The right hand expression is the solution of (6.5) with boundary values 1, and so is less than or equal to 1 in D. Thus we find

$$\int_D K(P, Q)\, dV \leq \delta^{-1};$$

this integral is uniformly bounded in $D + B$. We also note that $K(P, Q)$ is non-negative (2) in $D + B$. A calculation of the kind given in §3 now leads to the complete continuity of the operator based on $K(P, Q)$. Further details are here omitted.

The integral equation of the problem is

$$u(P) = \int_D N(P, Q)\, F(Q, u(Q))\, dV_Q + tv_0(P) + v_1(P), \tag{6.7}$$

where for $i = 0, 1$ we have

$$u_i(P) = \int_B N(P, q)\, g_i(q)\, dS_q. \tag{6.8}$$

Since $N(P, Q) \geq 0$ it follows from (6.4) that $v_1(P) > 0$, and we denote by v_1 and V_1 positive lower and upper bounds:

$$0 < v_1 \leq v_1(P) \leq V_1,$$

while similarly choosing bounds for $v_0(P)$:

$$v_0 \leq v_0(P) \leq V_0.$$

The operator T for this problem will now be defined as

$$\tilde{T}_k[u](P) = \int_D N(P, Q)\, F(Q, ku(Q))\, dV_Q + tv_0(P) + v_1(P), \tag{6.9}$$

while $t = \tilde{t}_k[u]$ is fixed by the condition

$$\max \tilde{T}_k[u] = M. \tag{6.10}$$

Setting

$$\Omega = \{u \mid -K < u < M + \epsilon\},$$

we find that for $u \in \Omega$, $F(P, u)$ satisfies an estimate

$$F(P, u) \leq A. \tag{6.11}$$

Then $t = \tilde{t}_k[u]$ has the bounds

$$\frac{|N_0A + V_1 - M|}{v_0} \leq t \leq \frac{M - v_1 + N_0F_0}{V_0}. \tag{6.12}$$

We therefore choose $-K$ less than $v_0^{-1}|N_0A + V_1 - M|$, which is possible since this quantity is independent of K. The equation

$$u = \tilde{T}_k[u]$$

now has no solutions on Ω' for $0 < k < 1$; and a unique solution for $k = 0$. The result now follows as before.

THEOREM V. *There exists a solution of* (6.1) *which satisfies the boundary condition* (6.2) *for some t, and has maximum value M.*

As in Theorem I, the right side of (6.3) could be replaced by a more general increasing function of t. Corresponding results for the Dirichlet and Robin boundary conditions and this differential equation can be established along the same lines of proof.

In conclusion we note that the uniqueness of solutions in all of these results has not been established.

References

1. S. Banach, *Théorie des opérations linéaires* (Warsaw, 1932).
2. S. Bergman and M. Schiffer, *Kernel functions in the theory of partial differential equations of elliptic type*, Duke Math. J., *15* (1948), 535–566.
3. ———, *Existence theorems for some quasilinear partial differential equations*, ONR report NRO 43068, (1952).
4. R. Courant and D. Hilbert, *Methoden der mathematischen Physik* (Berlin, Vol. I, 1931; Vol. II, 1937).
5. J. Leray and J. Schauder, *Topologie et équations fonctionnelles*, Ann. Ec. Norm. Sup., *51* (1934), 46–78.
6. L. Lichtenstein, *Vorlesungen über nichtlineare Integralgleichungen* (Berlin, 1931).

University of Toronto

SOME ALGEBRAIC PROPERTIES OF ASYMPTOTIC POWER SERIES

T. E. HULL

1. Introduction. Let us consider all power series of the form

$$c_0 + c_1z + c_2z^2 + \ldots + c_nz^n + \ldots .$$

It was shown first by Borel **(1)** that to each such series there corresponds a non-empty class of functions such that each function in the class has the given series as its asymptotic expansion about $z = 0$, the expansion being valid in a sector of the right half z-plane with vertex at the origin. Various generalizations of Borel's theorem have been given by Carleman **(1)**, van der Corput **(2)**, and Erdélyi **(3)**.

We shall be interested only in the case where the c_n are real and where z is a real, non-negative variable x. We are then led to the following special case of Borel's theorem. To any series

$$c_0 + c_1x + c_2x^2 + \ldots + c_nx^n + \ldots ,$$

there corresponds at least one function $f(x)$ such that

$$R_n(x)\, x^n = o(x^{n-1}), \quad x \to 0,$$

where

$$R_n(x)\, x^n = f(x) - c_0 - c_1x - c_2x^2 \ldots - c_{n-1}x^{n-1}, \quad n = 1, 2, 3, \ldots ,$$

is the remainder after n terms.

Because of Borel's theorem the expressions "asymptotic power series" and "formal power series" are equivalent; we shall refer to them as "asymptotic series" or simply as "series." We shall refer to the *class* of all sum functions $f(x)$ corresponding to a particular series as the asymptotic sum of the series.

It is obvious that the collection of all asymptotic series forms a ring under formal addition, subtraction, and multiplication and it is known **(4)** that this ring is isomorphic to the ring of all asymptotic sums.

It is the purpose of this paper to discuss, using primarily algebraic notions, some of the properties of these rings. To do so we pay particular attention to the fundamental role played by those special asymptotic series for which

(i) $c_0 > 0$,

(ii) there exists a sum function $f(x)$ such that, for all $x > 0$, $|R_n(x)| < |c_n|$ ($n = 0, 1, 2, \ldots$), where $R_0(x) = f(x)$ and otherwise $R_n(x)$ is defined as above.

Condition (ii) means that the remainder, with respect to $f(x)$, is numerically less than the first neglected term. Any series satisfying the properties (i) and

Received June 15, 1955. This research was supported by the United States Air Force, through the Office of Scientific Research of the Air Research and Development Command.

(ii) will be referred to as an S-series. Such series often arise in physical problems and because of their remainder property are especially useful in computations.

Our plan is to show first that the collection of all S-series is closed under formal addition and multiplication (but not subtraction). Since the distributive law will hold too, we shall call such a collection a semiring. Then we shall show that the full ring of all asymptotic series is generated from this semiring when we adjoin all differences to the semiring.

We may mention that such an imbedding of a semiring in a ring can arise in other contexts. The simplest of these is the imbedding of the semiring of all integers greater than or equal to some fixed non-negative number in the ring of all integers.

2. The S-series form a semiring.

We proceed now to prove the first of our two theorems.

THEOREM 1. *The S-series form a semiring. That is, the formal sum or product of two S-series is an S-series and the distributive law holds.*

We show first that the coefficients in an S-series must alternate in sign unless the series consists of only the constant term. Suppose that $c_n > 0$ $(n = 0, 1, 2, \ldots)$. Then, since

$$R_n(x) = c_n + R_{n+1}(x)\,x$$

and

$$|R_n(x)| < |c_n|,$$

we obtain

$$R_{n+1}(x)\,x < 0,$$

so that

$$R_{n+1}(x) < 0.$$

Moreover,

$$R_{n+1}(x) = c_{n+1} + R_{n+2}(x)\,x$$

and, letting $x \to 0$, we obtain

$$R_{n+1}(0+) = c_{n+1},$$

so that

$$c_{n+1} < 0.$$

Similarly, if $c_n < 0$, we obtain $c_{n+1} > 0$.

We have still to show that the coefficients must all be non-zero except in the special case where the series consists of only the constant term. We obviously cannot have any coefficient equal to zero unless the series terminates; but the series cannot terminate with the term c_nx^n $(n = 1, 2, 3, \ldots)$, because if it did we would have

$$R_{n-1}(x) = c_{n-1} + c_nx,$$

and x could always be chosen so large that

$$|R_{n-1}(x)| > |\epsilon_{n-1}|.$$

From now on we shall denote the non-constant S-series by, for example,

$$f^{\alpha} = \alpha_0 - \alpha_1 x + \alpha_2 x^2 - \ldots + (-1)^{n-1}\alpha_{n-1} x^{n-1} + (-1)^n R_n^{\alpha} x^n$$

and

$$f^{\beta} = \beta_0 - \beta_1 x + \beta_2 x^2 - \ldots + (-1)^{n-1}\beta_{n-1} x^{n-1} + (-1)^n R_n^{\beta} x^n,$$

where $\alpha_n, \beta_n > 0$ and $R_n^{\alpha}, R_n^{\beta} \geqslant 0.$

The Theorem requires us to show that the sum or product of two S-series is an S-series. The "sum" part of the proof is trivial. The "product" part is also trivial if one or both of the series is constant; for the other case we note that the remainder, with respect to $f^{\alpha} f^{\beta}$, after n terms in the formal product of the above two series can be written

$$(-1)^{n+1} (\alpha_0 R_n^{\beta} + \alpha_1 R_{n-1}^{\beta} + \ldots + \alpha_{n-1} R_1^{\beta} + R_n^{\alpha} R_0^{\beta}) x^{n+1}$$

while the $(n + 1)$th term in the formal product is

$$(-1)^{n+1} (\alpha_0\beta_n + \alpha_1\beta_{n-1} + \ldots + \alpha_n\beta_0) x^{n+1}.$$

The first of these two expressions is numerically less than or equal to the second so that condition (ii) is satisfied. The first term in the formal product is $\alpha_0\beta_0 > 0$ and so condition (i) is also satisfied. The product series is therefore an S-series with respect to the function $f^{\alpha} f^{\beta}$. It is obvious that the distributive law holds and so the Theorem is proven.

In fact we have shown that the semiring of all S-series is isomorphic to the semiring whose elements are the classes of sum functions which satisfy condition (ii). The semiring possesses a unit and a zero element which are simply the numbers 1 and 0 respectively.

It can also be shown that the formal substitution of an S-series in place of the variable in a convergent series produces another S-series, provided the coefficients of the convergent series are positive and its radius of convergence is greater than the constant term in the first S-series.

Incidentally, the non-constant S-series alone form a semiring without, of course, either a unit or a zero element. This semiring is an ideal, if differences are not allowed, in the larger semiring of all S-series.

3. The semiring generates the ring. We shall now show that the full ring of all asymptotic series is generated from the semiring of all S-series when we adjoin all differences to the semiring. The result can be formulated in the following way.

THEOREM 2. *Any asymptotic series can be written as the difference between two S-series.*

Suppose that $\alpha_0, \alpha_1 > 0$ and consider the series

$$\alpha_0 - \alpha_1 x + \alpha_2 x^2 - \ldots + (-1)^n \alpha_n x^n \ldots .$$

We shall show shortly that, if $\alpha_{n+1}/\alpha_n \to \infty$ as $n \to \infty$ and the inequalities

$$\alpha_{n+1}/\alpha_n > \alpha_n/\alpha_{n-1}, \qquad n = 1, 2, 3, \ldots,$$

are also satisfied, the series is an S-series. The proof of the Theorem is then straightforward; for, given any series

$$c_0 + c_1 x + c_2 x^2 + \ldots + c_n x^n + \ldots,$$

one can always choose some $\alpha_0, \alpha_1, \beta_0, \beta_1 > 0$ so that $\alpha_0 - \beta_0 = c_0$ and $\alpha_1 - \beta_1 = -c_1$. Then one can always choose pairs $\alpha_2, \beta_2 > 0$, $\alpha_3, \beta_3 > 0, \ldots$ in turn so that α_{n+1}/α_n and $\beta_{n+1}/\beta_n \to \infty$ as $n \to \infty$, and

$$\frac{\alpha_{n+1}}{\alpha_n} > \frac{\alpha_n}{\alpha_{n-1}}, \qquad \frac{\beta_{n+1}}{\beta_n} > \frac{\beta_n}{\beta_{n-1}}, \qquad n = 1, 2, 3, \ldots$$

and so that $\alpha_n - \beta_n = (-1)^n c_n$. The α-series and the β-series so formed are then both S-series and their difference is the given series. The Theorem is then proven.

We have only to show that the conditions assumed for α_n in the above paragraph ensure that the corresponding α-series is an S-series. For a series to be an S-series, conditions (i) and (ii) must be satisfied. Condition (i) is satisfied since we have assumed that $\alpha_0 > 0$; in fact our assumptions guarantee that all $\alpha_n > 0$. We can show that condition (ii) is also satisfied by constructing the required "sum" function.

We define the intervals I_n in the following way: I_0 is the interval $0 < x < \infty$ and I_n $(n = 1, 2, 3, \ldots)$ is the interval $0 < x < \alpha_{n-1}/\alpha_n$. Putting

$$\mu_n(x) = \begin{cases} 1, & x \in I_n, \\ 0, & x \notin I_n, \end{cases}$$

we define

$$f(x) = \sum_{i=0}^{\infty} (-1)^i \mu_i(x)\, \alpha_i x^i.$$

This series converges for all x—in fact, it terminates for each x. Therefore $f(x)$ is defined.

For our purposes the essential points are the following. For each x, the terms which appear in the series for $f(x)$ *decrease* in magnitude with increasing subscript (unless only the first term appears). The terms which do not appear in the series for $f(x)$ are *non-decreasing* in magnitude with increasing subscript.

Then, if we suppose that $x \in I_N - I_{N+1}$ $(N = 0, 1, 2, \ldots)$ and if we take account of the fact that all terms are alternating in sign, we can easily pick

out one term which dominates $R_{n+1}(x)\, x^{n+1}$ ($n = -1, 0, 1, \ldots$). We obtain

$$|R_{n+1}(x)\, x^{n+1}| \begin{cases} < \alpha_{n+1}\, x^{n+1}, & n < N, \\ = 0, & n = N, \\ < \alpha_n x^n, & n > N. \end{cases}$$

The last expression is, in turn, $< \alpha_{n+1}\, x^{n+1}$ when $n > N$, so that condition (ii) is satisfied. (If $N = \infty$ we of course need to consider only the case where $n < N$, and if $N = 0$ we need to consider only the cases where $n > N$.)

The conditions assumed for α_n are not necessary for the series to be an S-series; this can be seen by considering the expansion of e^{-x}. Moreover it can be shown that the S-series which do satisfy these conditions are closed under addition but not under multiplication.

Incidentally we have in fact shown that the semiring of non-constant S-series also generates the full ring with the adjoining of all differences.

4. Concluding remarks. D. C. Murdoch has pointed out that the above results enable one to define a partial ordering on the ring of all power series. One series $a(x)$ can be defined to be "greater than or equal to" another series $b(x)$ if and only if their difference is an S-series. By using a procedure analogous to that used in the first part of Theorem 2, one can then always construct an upper bound and a lower bound to any pair of series. However, it is also possible to show that neither the least upper bound nor the greatest lower bound required for a lattice can exist.

Algebraic and other properties of asymptotic series have been considered by Popken **(5)**. He discusses the ring of all asymptotically finite functions (and so does not restrict his attention to power series) and he shows, for example, that this ring is complete with respect to a certain non-Archimedean pseudo-valuation.

We wish to thank B. N. Moyls for many interesting discussions during the preparation of this paper, and also the referee, particularly for pointing out an error in the proof of Theorem 2.

References

1. T. Carleman, *Les fonctions quasianalytiques* (Paris, 1926).
2. J. G. van der Corput, *Asymptotic expansions I. Fundamental theorems of asymptotics* (Univ. of California, Berkeley, 1954).
3. A. Erdélyi, *Asymptotic expansions* (Cal. Tech., Pasadena, 1955).
4. H. Poincaré, *Sur les intégrales irregulières des équations linéaires*, Acta Math., *8* (1886) 295–344.
5. J. Popken, *Asymptotic expansions from an algebraic standpoint*, Nederl. Akad. Wetensch. Proc., Ser. A., *56* (Indagationes Math., *15* (1953), 131–143).

University of British Columbia

ASYMPTOTIC EXPANSIONS

LEO MOSER AND MAX WYMAN

1. Introduction. Let $a_1 a_2, \ldots, a_m$ be a set of real non-negative numbers and let

1.1 $$P(x) = a_1x + a_2x^2 + \ldots + a_mx^m \qquad (a_m \neq 0).$$

Many combinatorial problems can be reduced to the study of numbers B_n generated by

1.2 $$\sum_{n=0}^{\infty} B_n x^n/n! = e^{P(x)}.$$

Some problems of this type were treated by Touchard **(7)**, Jacobsthal **(3)**, Chowla, Herstein, Moore and Scott **(1; 2)**, and the present authors **(4)**. In **(2)**, the problem of finding asymptotic formulae for B_n in terms of $P(x)$ was proposed. Essentially the same problem was solved earlier by Pólya **(5)**, as a by-product of an investigation of the zeros of the derivatives of certain functions. The object of the paper is to give a different and more explicit solution to this problem. Furthermore our method yields complete asymptotic expansions, while that of Pólya gave only the first term.

2. Preliminary notions. Since some of the coefficients a_k in (1.1) may be zero, $P(x)$ will in general have the form

2.1 $$P(x) = b_1x^r + b_2x^s + \ldots + a_mx^m,$$

where the coefficients $b_1, b_2, \ldots, a_m$ are positive. In what follows we shall assume

2.2 $$(r, s, \ldots, m) = 1.$$

This involves no essential loss of generality since one can always reduce the problem to this case by a substitution of the form $y = x^q$.

LEMMA 1. *If* $0 < \theta < \pi$ *and* $\cos r\theta = 1$, $\cos s\theta = 1, \ldots, \cos m\theta = 1$ *then* $\theta = 0$.

Proof. If $0 < \theta < \pi$ then $\cos r\theta = 1$ implies the existence of positive integers a and b, $(a, b) = 1$, $b > 1$, such that $\theta = \pi\, a/b$. Now $r\theta = r\pi\, a/b$, $s\theta = s\pi\, a/b, \ldots, m\theta = m\pi\, a/b$ are each integral multiples of 2π. Hence b divides $r, s, \ldots, m$, which contradicts (2.2).

Corresponding to $P(x)$ as defined in (1.1) we define a trigonometric polynomial $S(R, \theta)$ by

Received July 28, 1955.

2.3 $$S(R, \theta) = \tfrac{1}{2}[P(Re^{i\theta}) + P(Re^{-i\theta})] = \sum_{k=1}^{m} a_k R^k \cos k\,\theta.$$

Further we define ϵ by

2.4 $$\epsilon = R^{(1-4m)/8},$$

and prove

LEMMA 2. *For $\epsilon < \theta < \pi$ and R sufficiently large, $S(R, \theta) < S(R, \epsilon)$.*

Proof. Since $a_k > 0$ for $k = 1, 2, \ldots, m$. $S(R, \theta)$ assumes its greatest value at $\theta = 0$. Also by (1.1), (2.3) and (2.4) we have

2.5 $$S(R, 0) - S(R, \epsilon) = \sum_{k=1}^{m} a_k R^k (1 - \cos k\epsilon) = O(R^{\frac{1}{4}}).$$

On the other hand, there exists by (2.2) and Lemma 1, a positive integer $t < m$ such that $\cos(t\theta) \neq 1$ and $a_t \neq 0$. Hence for fixed θ,

2.6 $$S(R, 0) - S(R, \theta) = \sum_{k=1}^{m} a_k R^k (1 - \cos k\,\theta) > C_1 R^t,$$

where C_1 is a fixed positive constant. Comparing (2.5) and (2.6) gives the required result.

3. Asymptotic formulae. By (1.2) and Cauchy's theorem

3.1 $$B_n = \frac{n!}{2\pi i} \int_c \frac{e^{P(z)}}{z^{n+1}}\, dz,$$

where c denotes the circle $z = Re^{i\theta}$. We note that R, the radius of the circle, is arbitrary. From (3.1) we obtain

3.2 $$B_n = A \int_{-\pi}^{\pi} e^{F(R,\theta)} d\theta,$$

where

3.3 $$A = n!\, e^{P(R)}/2\pi R^n,$$

and

3.4 $$F(R, \theta) = P(Re^{i\theta}) - P(R) - in\theta.$$

Let I be defined by

3.5 $$I = \int_{\epsilon}^{\pi} e^{F(R,\theta)} d\theta,$$

where ϵ is given by (2.4).

LEMMA 3. $|I| = O(\exp(-R^{\frac{1}{4}}))$.

Proof. Clearly

$$|I| < \int_{\epsilon}^{\pi} e^{S(R,\theta) - S(R,0)} d\theta,$$

and the required result follows from (2.5) and Lemma 2.

Since we will show that the integral in (3.2) can be expanded in powers of $1/R$, we may neglect integrals of type (3.5) and write

3.6 $$B_n \sim A \int_{-\epsilon}^{\epsilon} e^{F(R,\theta)} d\theta.$$

Our next step is to expand $F(R, \theta)$ in a Maclaurin series of the form

3.7 $$F(R, \theta) = \sum_{j=1}^{\infty} C_j(R)(i\theta)^j/j!,$$

where

3.8 $$C_1(R) = \sum_{k=1}^{m} k\, a_k R^k - n,$$

and

3.9 $$C_j(R) = \sum_{k=1}^{m} k^j a_k R^k \qquad (j > 1).$$

At this stage we choose R so that

3.10 $$C_1(R) = 0.$$

For large n, (3.10) will have a unique solution which may be calculated by iteration starting with

3.11 $$R \sim (n/m\, a_m)^{1/m}.$$

When (3.10) holds, (3.7) can be written in the form

3.12 $$F(R, \theta) = -\tfrac{1}{2} C_2(R)\theta^2 + \sum_{j=3}^{\infty} C_j(R)(i\theta)^j/j!\,.$$

In order to simplify some of the expressions which occur, we introduce the following notations:

3.13 $$R = z^{-2}, \quad z = R^{-\frac{1}{2}},$$

3.14 $$\bar{C}_j(z) = z^{2m} C_j(z^{-2}) = \sum_{k=1}^{m} k^j a_k z^{2m-2k}, \qquad (j > 1),$$

3.15 $$f_j(z) = z^{m(j-2)}\, \bar{C}_j(z)\{2/\bar{C}_2(z)\}^{\frac{1}{2}j},$$

3.16 $$\lambda = \epsilon(C_2(R)/2)^{\frac{1}{2}},$$

3.17 $$\phi = \theta(C_2(R)/2)^{\frac{1}{2}},$$

3.18 $$H = A(2/C_2(R))^{\frac{1}{2}},$$

3.19 $$\psi(z, \phi) = \sum_{j=3}^{\infty} f_j(z)(i\phi)^j/j!\,.$$

If we now make the substitution (3.17) in (3.6) and use (3.12) and (3.19) we obtain

3.20 $$B_n \sim H \int_{-\lambda}^{\lambda} e^{-\phi^2 + \psi(z,\phi)} d\phi.$$

From (2.4), (3.9) and (3.16) we see that for R large,

3.21 $$K_2R^{1/8} > \lambda > K_3R^{1/8},$$

where K_2 and K_3 are fixed positive constants. Further, for R sufficiently large it is not difficult to show that there exists an interval $-\sigma < z < \sigma$ for which $\psi(z, \phi)$ and $e^{\psi(z,\phi)}$ have Maclaurin expansions in z of the form

3.22 $$\psi(z, \phi) = \sum_{k=1}^{\infty} \psi_k(\phi)\, z^k$$

and

3.23 $$e^{\psi(z,\phi)} = \sum_{r=0}^{\infty} \Psi_r(\phi)\, z^r, \quad \Psi_0(\phi) = 1,$$

which are uniformly convergent for $|\phi| < \lambda$. It is further easy to justify the fact that the $\psi_k(\phi)$ are given by means of (3.19) to be

3.24 $$\psi_k(\phi) = \sum_{j=3}^{\infty} \frac{1}{k!}\left[\frac{d^k f_j(z)}{dz^k}\right]_{z=0} \frac{(i\phi)^j}{j!}.$$

From (3.19) we see that $\psi_k(\phi)$ are polynomials in ϕ and hence $\Psi_k(\phi)$ are also polynomials in ϕ. In fact $\Psi_{2n}(\phi)$ contains only even powers of ϕ while $\Psi_{2n+1}(\phi)$ only contains odd powers of ϕ.

Using (3.24) in (3.20) we have

3.25 $$B_n \sim H\left[\sum_{k=0}^{s-1}\left(\int_{-\lambda}^{\lambda} \Psi_k(\phi)\, e^{-\phi^2} d\phi\right) z^k + {}'R_s\right],$$

where

3.26 $$R_s = \int_{-\lambda}^{\lambda} e^{-\phi^2} \sum_{k=s}^{\infty} \Psi_k(\phi)\, z^k d\phi\,.$$

Using (3.21) and the fact that the $\Psi_k(\phi)$ are polynomials in ϕ, (3.25) yields

3.27 $$B_n \sim H\left[\sum_{k=0}^{s-1}\left(\int_{-\infty}^{\infty} \Psi_k(\phi)\, e^{-\phi^2} d\phi\right) z^k + R_s\right],$$

where R_s is still given by (3.26). In order to complete our proof it remains to show that for fixed s, $R_s = O(z^s)$. If this is so our complete asymptotic formula becomes

3.28 $$B_n \sim H\left[\sum_{k=0}^{\infty} \int_{-\infty}^{\infty} \Psi_k(\phi)\, e^{-\phi^2} d\phi / R^{\frac{1}{2}k}\right].$$

Finally, in view of the remarks following (3.24), the integrals in (3.28) will vanish for odd k and (3.28) can be put in the form

3.29 $$B_n \sim H\left[\sum_{k=0}^{\infty} \left(\int_{-\infty}^{\infty} \Psi_{2k}(\phi)\, e^{-\phi^2} d\phi\right)/R^k\right].$$

We shall now consider $\psi(z, \phi)$ as given by (3.19) to be a function of a com-

plex variable z and a real variable ϕ. We restrict z to be in a neighborhood $|z| < \sigma < 1$ and ϕ to be bounded. Under these restrictions (3.14) yields

3.30 $$|\tfrac{1}{2}\bar{C}_2(z)| > \left| m^2 a_m - \sum_{k=1}^{m-1} k^2 a_k \sigma^{2m-2k} \right| .$$

Clearly, by taking σ small enough we may say that

3.31 $$|\tfrac{1}{2}\bar{C}_2(z)| > \alpha^2,$$

where α is a positive constant.

Similarly from (3.19) and the fact that $|z| < 1$ we can say that

3.32 $$|\bar{C}_j(z)| < \sum_{k=1}^{m} k^j |a_k| < a(m^{k+1}),$$

where $a = \max(|a_1|, |a_2|, \ldots, |a_m|)$. Hence from (3.15), (3.31) and (3.32) we obtain

3.33 $$|f_j(z)| < a(m^{j+1})/\alpha^j,$$

where a, m, and α are independent of j and z. From (3.3) and Cauchy's theorem on derivatives we obtain

3.34 $$\left| \frac{d^k f_j(z)}{dz^k} \right|_{z=0} < a(m^{j+1})\, k!/\alpha^j \sigma^k.$$

By introducing the notation $am = M$, $m/\alpha = K$, $\sigma = 1/S$, (3.34) may be written

3.35 $$\left| \frac{d^k f_j(z)}{dz^k} \right|_{z=0} < M K^j k!\, S^k.$$

From (3.15) the derivatives in (3.35) vanish for $m(j - 2) > k$. Since m is a positive integer, $k + 2 > (k/m) + 2$. Hence (3.24) can be written

3.36 $$\psi_k(\phi) = \sum_{j=3}^{k+2} \frac{1}{k!} \left[\frac{d^k f_j(z)}{dz^k} \right]_{z=0} \frac{(i\phi)^j}{j!} .$$

From (3.35) we now obtain

3.37 $$|\psi_k(\phi)| < MS^k \sum_{j=3}^{k+2} (K|\phi|)^j/j!,$$

and by induction on k we easily deduce

3.38 $$|\psi_k(\phi)| < MS^k\, [K|\phi|]^2\, [1 + K|\phi|]^k.$$

By a lemma proved in **(4)**, (3.38) implies

3.39 $$|\Psi_j(\phi)| < M[K|\phi|]^2[1 + M(K\phi)^2]^{j-1} S^j (1 + K|\phi|)^j.$$

From (3.39) we obtain

3.40 $$\left| \sum_{j=s}^{\infty} \Psi_j(\phi)\, z^j \right| < \frac{1}{T} M[K|\phi|]^2[1 + M(K\phi)^2]^{s-1} S^s (1 + K|\phi|)^s |z|^s$$

where T is given by

3.41 $$T = 1 - [1 + M(K|\phi|)^2][1 + K|\phi|]|z|.$$

We now revert to real values of z. Recalling that $z = R^{-\frac{1}{2}}$ and $|\phi| < \lambda$ we have, from (3.21), that $\lambda < K_2R^{1/8}$,

3.42 $$|\phi|^3|z| = O(R^{-1/8}).$$

Hence for R sufficiently large we have $T > \frac{1}{2}$. Thus (3.40) yields

3.43 $$\left|\sum_{j=s}^{\infty} \Psi_j(\phi)\, z^j\right| < Q_s(|\phi|)\, z^s,$$

where $Q_s(|\phi|)$ is a polynomial in $|\phi|$. From (3.26) we now obtain

3.44 $$|R_s| < \int_{-\lambda}^{\lambda} e^{-\phi^2} Q_s(|\phi|)\, d\phi\, z^s < z^s \int_{-\infty}^{\infty} e^{-\phi^2} Q_s(|\phi|)\, d\phi.$$

Since $Q_s(|\phi|)$ is a polynomial, the last integral of (3.44) exists and hence

3.45 $$R_s = O(z^s).$$

This completes the proof of the main result (3.29) which can be written in the form

3.46 $$B_n \sim \frac{n!\, e^{P(R)}}{2\pi R^n}\left[\frac{2}{C_2(R)}\right]^{\frac{1}{2}} \sum_{k=0}^{\infty}\left(\int_{-\infty}^{\infty} \Psi_k(\phi)\, e^{-\phi^2} d\phi / R^{k}\right),$$

where R is determined by

3.47 $$\sum_{k=1}^{m} k\, a_k R^k - n = 0.$$

In concluding this section we might point out that $\Psi_0(\phi) = 1$. Hence the first term of the asymptotic expansion is easily calculated. If we introduce the operator Θ by

3.48 $$\Theta = R\frac{d}{dR},$$

then the first term of the expansion is given by

3.49 $$B_n \sim \frac{n!\, e^{P(R)}}{R^n}\left[\frac{1}{2\pi\Theta^2 P(R)}\right]^{\frac{1}{2}},$$

and R as a function of n is given by

3.50 $$\Theta P(R) = n.$$

4. Applications. To illustrate applications of the method we consider three special cases.

Example 1. $P(x) = x$.

In this case the numbers B_n are all 1. However, since the asymptotic formula obtained by our method involves the factor $n!$, it will lead in this case to Stirling's expansion for $n!$. Equations (3.8), (3.9) and (3.10) yield $R = n$ and $C_j(R) = R = n$ $(j > 1)$. Applying (3.46) we obtain

4.1 $$1 \sim \frac{n!\,e^n}{(2\pi)^{\frac{1}{2}} n^{n+\frac{1}{2}}}\left(1 - \frac{1}{12n} + \dots\right)$$

or

4.2 $$n! \sim \left(\frac{n}{e}\right)^n (2\pi n)^{\frac{1}{2}}\left(1 + \frac{1}{12n} + \dots\right)$$

as required.

Example 2. $P(x) = x + (x^p/p)$.

In this case it is known **(3)** that if p is a prime then $B_n = B_{n,p}$ is the number of solutions of $x^p = 1$ in the symmetric group of degree n. The case $p = 2$ was treated in **(1)** and **(4)** and the result for $p > 2$ was announced in **(4)**. In this case we have

4.3 $$P(R) = R + (R^p/p),$$

4.4 $$C_1(R) = R + R^p - n = 0$$

and

4.5 $$C_2(R) = R + pR^p.$$

From these and (3.49) the first term of the asymptotic expansion is given by

4.6 $$B_{n,p} \sim \frac{n!\exp(R + R^p/p)}{R^n[2\pi(R + pR^p)]^{\frac{1}{2}}}.$$

Now using (4.2) and (4.4) we obtain,

4.7 $$n! \sim n^n e^{-n}(2\pi n)^{\frac{1}{2}},$$

4.8 $$\exp\left(\frac{R^p}{p}\right) = \exp\left(\frac{n}{p} - \frac{R}{p}\right).$$

Also

4.9 $$R^n = (n - R)^{n/p} = n^{n/p}\exp\left(\frac{n}{p}\log\left(1 - \frac{R}{n}\right)\right).$$

Expanding $\log(1 - R/n)$ yields

4.10 $$R^n \sim n^{n/p}\exp\left(-\frac{R}{p} - \frac{R^2}{2pn}\right)$$

Finally

4.11 $$(R + pR^p)^{\frac{1}{2}} \sim (pn)^{\frac{1}{2}}.$$

Using (4.7) to (4.11) in (4.6) yields

4.12 $$B_{n,p} \sim \left(\frac{n}{e}\right)^{n(1-1/p)} p^{-\frac{1}{2}}\exp\left(R + \frac{R^2}{2pn}\right).$$

We now consider two cases:

Case 1. $p = 2$. Here $e^{R+(R^2/2np)} \sim e^{n^{\frac{1}{2}}-\frac{1}{2}+\frac{1}{4}} = \exp(n^{\frac{1}{2}} - \frac{1}{4})$.

Case 2. $p > 2$. Here $e^{R+(R^2/2np)} \sim \exp(n^{1/p})$.
Thus we obtain

4.13 $$B_{n,2} \sim \left(\frac{n}{e}\right)^{\frac{1}{2}n} \exp(n^{\frac{1}{2}})\, 2^{-\frac{1}{2}} e^{-\frac{1}{4}}$$

and

4.14 $$B_{n,p} \sim \left(\frac{n}{e}\right)^{n(1-1/p)} p^{-\frac{1}{2}} \exp(n^{1/p}) \qquad (p > 2).$$

Example 3. $P(x) = 2tx + x^2 \qquad (t > 0)$.

Here $B_n = B_n(t)$ are polynomials in t. In this case we have

4.15 $$P(R) = 2tR + R^2,$$

4.16 $$\Theta P(R) = 2tR + 2R^2 = n,$$

4.17 $$R = \tfrac{1}{2}[-t + (2n + t^2)^{\frac{1}{2}}].$$

From these and (3.49) we obtain

4.18 $$B_n(t) \sim \frac{n! \exp(2Rt + R^2)}{R^n} \frac{1}{[2\pi(2Rt + 4R^2)]^{\frac{1}{2}}},$$

where R is given by (4.17).

By computing the first two terms of the asymptotic expansion, $B_n(t)$ can be put in the form

4.19 $$B_n(t) \sim \left(\frac{n}{e}\right)^{\frac{1}{2}n} 2^{\frac{1}{2}(n-1)} \exp((2n)^{\frac{1}{2}}t - \tfrac{1}{4}t^2)\left\{1 + \frac{t^3 + 2t}{6(2n)^{\frac{1}{2}}}\right\}.$$

Our method restricts t to be positive. However, the above result is valid also for $t < 0$. $B_n(t)$ is of course related to the Hermite polynomials and (4.19) can be checked by means of the known expansion formula for these polynomials given in (6, p. 194).

5. Conclusion. We have given here a method of finding asymptotic expansions for numbers or functions whose generating function is of the form $e^{P(x)}$. In this paper we have restricted $P(x)$ to be a polynomial in x with non-negative coefficients. If this severe restriction on $P(x)$ is relaxed (3.49) may no longer be valid. We hope, in a subsequent paper, to show how the method may be modified to cope with the case of less restricted functions $P(x)$.

REFERENCES

1. S. Chowla, I. N. Herstein, and K. Moore, *On recursions connected with symmetric groups* I, Can. J. Math., *3* (1951), 328–334.
2. S. Chowla, I. N. Herstein, and W. R. Scott, *The solutions of $x^d = 1$ in symmetric groups*, Norske Vid. Selsk., *25* (1952), 29–31.
3. E. Jabobsthal, *Sur le nombre d'éléments du group symmetrique S_n dont l'ordre est un nombre premier*, Norske Vid. Selsk., *21* (1949), 49–51.
4. L. Moser and M. Wyman, *On solutions of $x^d = 1$ in symmetric groups*, Can. J. Math., 7 (1955), 159–168.
5. G. Pólya, *Ueber die Nullstellen sukzessiver Derivierten*, Math. Z., *12* (1922), 36–60.
6. G. Szegö, *Orthogonal Polynomials*, Amer. Math. Soc. Coll. Publications (New York, 1939).
7. J. Touchard, *Sur les cycles des substitutions*, Acta Math., *70* (1939), 242-297.

University of Alberta

THE ASYMPTOTIC SERIES FOR A CERTAIN CLASS OF PERMUTATION PROBLEMS

N. S. MENDELSOHN

1. Introduction. This paper is concerned with problems connected with the permutations of the integers $1, 2, \ldots, n$ subject to certain special restrictions. One such class of problems, the so-called "card matching" problems, deals with conditions of the type, "the number i is in the jth position," "the number k is in the mth position," etc. The given conditions need not be compatible, i.e. a meaningful problem results from having amongst the set of conditions such conditions as "1 is second," "2 is second," "2 is third." In a permutation of $1, 2, 3, \ldots, n$ and a set of conditions S we will say that there are r "hits" if exactly r of the given conditions are fulfilled. Amongst the $n!$ permutations of the numbers $1, 2, 3, \ldots, n$, suppose there are $N(r)$ in which there are r hits. The problem of determining $N(r)$ has been treated in **(3**; **6**; **7)**. These results may be expressed in the language of probability by saying that $M(r) = N(r)/n!$ is the probability of exactly r hits.

A second type of problem deals with the so-called relative conditions, i.e., conditions such as "i immediately precedes j." These problems are dealt with in much the same way as the previous type in **(3)**, and for the purposes of this paper will not require a separate treatment.

For a fairly large class of problems of both types the distribution of $M(r)$ is asymptotic to a Poisson distribution. In fact, in these cases, it is possible to write $M(r)$ in the form:

$$M(r) = e^{-A}\frac{A^r}{r!}\left(1 + \frac{c_1}{n} + \frac{c_2}{n(n-1)} + \frac{c_3}{n(n-1)(n-2)} + \ldots\right). \tag{1}$$

In general the c_i are polynomials in r of degree at most $2i$. It is the purpose of this paper to show how to compute $A, c_1, c_2, \ldots$. The determination of A, $c_1, c_2, \ldots$, does not require a knowledge of the exact expression for $N(r)$. It suffices to have a difference equation for a certain polynominal operator associated with the given set of conditions. The computation can be carried out completely from a knowledge of the coefficients which appear in the difference equation, together with the initial conditions necessary to fix the solution of the difference equation.

2. The general problem. In what follows, the discussion will be confined to the card-matching type of problem although one of the illustrations given in the end will deal with a "relative condition" problem. If p_{ij} denotes the

Received July 22, 1955.

condition "i is in position j" and $N(0)$ and $N(r)$ denote the number of permutations of $1, 2, \ldots, n$ in which there are 0 and r hits respectively, the method of inclusion and exclusion yields the following formulae for $N(0)$ and $N(r)$:

$$N(0) = \sum_{k=0}^{n} (-1)^k \phi_{n,k}(n-k)!$$

and

$$N(r) = \sum_{k=r}^{n} (-1)^{k+r} \phi_{n,k} \binom{k}{r} (n-k)!\,.$$

In these formulae ϕ_{nk} represents the number of ways in which exactly k compatible conditions may be chosen from the set of all p_{ij}.

If $M(0)$ and $M(r)$ are the probabilities of 0 and r hits respectively, the relevant formulae are:

$$M(0) = \sum_{k=0}^{n} (-1)^k \phi_{n,k} \frac{(n-k)!}{n!}$$

and

$$M(r) = \sum_{k=r}^{n} (-1)^{k+r} \phi_{n,k} \binom{k}{r} \frac{(n-k)!}{n!}.$$

If $\psi(t)$ is the generating function of the number of hits, i.e.

$$\psi(t) = \sum_{r=0}^{n} M(r)\, t^r,$$

we have:

$$\psi(t) = \sum_{r=0}^{n} \sum_{k=r}^{n} (-1)^{k+r} \phi_{n,k} \binom{k}{r} \frac{(n-k)!}{n!}\, t^r$$

$$= \sum_{k=0}^{n} \phi_{n,k} (t-1)^k \frac{(n-k)!}{n!}.$$

The determination of $\phi_{n,k}$ has been treated in **(3; 6; 7)**. Perhaps the most interesting representation has been given by Kaplansky and Riordan in **(6)**. In this representation, for each condition p_{ij} the cell in the ith row, jth column in an $(n \times n)$ chessboard is marked. It is easily seen that $\phi_{n,k}$ is the number of ways of putting k non-attacking rooks on the marked squares of the board. This representation makes it easy in special cases to obtain explicit formulae for $\phi_{n,k}$, and in more complicated cases it simplifies the determination of recurrence relationships.

Fréchet **(2)** gives a thorough discussion of the method of inclusion and exclusion on which the formulae of this section are based.

3. The symbolic representation. By the use of the difference operator E, defined as $Ef(n) = f(n+1)$ the formulae for $N(0)$, $N(r)$, $M(0)$ $M(r)$ may be expressed in the forms:

$$N(0) = P_n(E)\, f(0),$$
$$N(r) = P_n(E)\, g_r(0),$$
$$M(0) = P_n(E)\, f^*(0),$$
$$M(r) = P_n(E)\, g_r^*(0),$$

where

$$P_n(E) = \sum_{k=0}^{n} (-1)^k \phi_{n,k} E^k,$$
$$f(t) = (n-t)!$$
$$g_r(t) = \begin{cases} (-1)^r \binom{t}{r} (n-t)!, & t \geqslant r, \\ 0, & t < r, \end{cases}$$
$$f^*(t) = \frac{(n-t)!}{n!},$$
$$g_r^*(t) = \begin{cases} (-1)^r \binom{t}{r} \dfrac{(n-t)!}{n!}, & t \geqslant r, \\ 0, & t < r. \end{cases}$$

In **(3; 7)** methods of obtaining difference equations for $P_n(E)$ are given and in a number of cases these lead to explicit formulae. This paper is concerned mostly with a determination of the asymptotic series for $M(0)$ and $M(r)$ in the cases where the difference equation for $P_n(E)$ is of a special form. In a large class of problems discussed in the literature the polynomial $P_n(E)$ does indeed have a difference equation of the required form.

4. Some illustrative examples. A number of examples (mostly classical) are given here to illustrate in concrete terms the type of problem with which this discussion is concerned. These examples have also served to verify the correctness of formulae which are developed later in this paper. Such verification is necessary since the computations are quite formidable.

Example 1. *Problème des rencontres.* In this example the set of conditions are: for $i = 1, 2, \ldots, n$, "i is in position i." In this case the formulae become:

$$M(0) = 1 - \frac{1}{1!} + \frac{1}{2!} - \ldots + (-1)^n \frac{1}{n!},$$
$$M(r) = \frac{1}{r!}\left(1 - \frac{1}{1!} + \frac{1}{2!} - \ldots + (-1)^{n-r} \frac{1}{(n-r)!}\right),$$
$$P_n(E) = (1-E)^n.$$

$P_n(E)$ satisfies the recurrence formula

$$P_n(E) = (1-E)\, P_{n-1}(E).$$

Asymptotic formulae are:

$$M(0) \sim e^{-1}, \quad M(r) \sim \frac{e^{-1}}{r!}.$$

In this case the general asymptotic series of the form

$$M(r) = \frac{e^{-A}A^r}{r!}\left(1 + \frac{c_1}{n} + \frac{c_2}{n(n-1)} + \dots\right)$$

does not exist. The reason for this is that convergence to the asymptotic value is so rapid that c_1, c_2, etc. are all 0.

Example 2. *Problème des ménages.* The conditions of this problem are: for $i = 1, 2, \dots, n - 1$, "i is in ith position" and "i is in $(i + 1)$th position," together with "n is in nth position" and "n is in first position." The requisite formulae are:

$$M(0) = \sum_{i=0}^{n} (-1)^i \binom{2n}{2n-i}\binom{2n-i}{i}\frac{(n-i)!}{n!},$$

$$M(r) = \sum_{i=r}^{n} (-1)^{r+i} \binom{2n}{2n-i}\binom{2n-i}{i}\binom{i}{r}\frac{(n-i)!}{n!};$$

the recurrence formula for $P_n(E)$ is given by,

$$P_n(E) = (1 - 2E)\, P_{n-1}(E) - E^2 P_{n-2}(E),$$

and the asymptotic formulae are:

$$M(0) = e^{-2}\left(1 - \frac{1}{(n-1)} + \frac{1}{2!(n-1)_2} + \dots + \frac{(-1)^i}{i!(n-1)_i} + \dots\right)$$

$$M(r) = \frac{e^{-2}2^r}{r!}\left(1 - \frac{(r-1)(r-4)}{4n} + \frac{r^4 - 14r^3 + 51r^2 - 38r - 16}{32n(n-1)}\right) + O(n^{-3}).$$

Here n_k is the Jordan factorial notation for $n(n - 1)(n - 2) \dots (n - k + 1)$. These results were obtained by Kaplansky and Riordan in **(5)** and check with the formula developed here.

Example 3. *Ménages non-circulaires.* This differs from Example 2 only in the omission of the condition "n is in first position." Here the formulae are:

$$M(0) = \sum_{i=0}^{n} (-1)^i \binom{2n-i}{i}\frac{(n-i)!}{n!},$$

$$M(r) = \sum_{i=r}^{n} (-1)^{r+i} \binom{2n-i}{i}\binom{i}{r}\frac{(n-i)!}{n!};$$

the recurrence formula for $P_n(E)$ is,

$$P_n(E) = (1-2E)\, P_{n-1}(E) - E^2\, P_{n-2}(E),$$

which is identical with that of the ordinary ménages problem; the asymptotic formulae are

$$M(0) = e^{-2}\left(1 - \frac{1}{2n(n-1)}\right) + O(n^{-3}),$$

and

$$M(r) = \frac{e^{-2}2^r}{r!}\left(1 - \frac{r(r-3)}{4n} + \frac{r^4 - 10r^3 + 23r^2 + 2r - 16}{32n(n-1)}\right) + O(n^{-3}).$$

These results were also given in **(5)**.

Example 4. *Rook-king problem.* This example is a case of a relative condition problem. The conditions are "1 immediately precedes 2," "n immediately precedes $n - 1$" and for $i = 2, 3, 4, \ldots, n - 1$, "$i$ immediately precedes $(i - 1)$" and "i immediately precedes $(i + 1)$." This problem had been treated in **(3)** and **(4)**. No convenient exact expressions for $M(0)$ and $M(r)$ have been found. The recurrence formula for $P_n(E)$ is given by

$$P_n(E) = (1 - E)\, P_{n-1}(E) - EP_{n-2}(E).$$

The formulae developed later yield the following asymptotic series:

$$M(0) = e^{-2}\left(1 - \frac{2}{n(n-1)}\right) + O(n^{-3}),$$

and

$$M(r) = \frac{e^{-2}2^r}{r!}\left(1 + \frac{r(3-r)}{2n} + \frac{r^4 - 8r^3 + 9r^2 + 22r - 16}{8n(n-1)}\right) + O(n^{-3}).$$

Example 5. The final example is one which has not appeared anywhere in the literature. The set of conditions is: "1 is 2nd," "n is $(n - 1)$th" and for $i = 2, 3, 4, \ldots, n - 1$, "$i$ in $(i - 1)$th" and "i is $(i + 1)$th." In terms of the chessboard representations the marked squares are precisely those on the two diagonals adjacent to the main diagonal. The recurrence formula for $P_n(E)$ has been obtained by the present author by the method given in his paper **(7)**. Exact formulae for $M(0)$ and $M(r)$ are not readily obtained but $P_n(E)$ satisfies the recurrence formula

$$P_n(E) = (1 - E)\, P_{n-1}(E) + (-E + E^2)\, P_{n-2}(E) + E^3P_{n-3}(E).$$

This together with

$$P_2(E) = 1 - 2E + E^2, \quad P_3(E) = (1 - 4E + 4E^2),$$
$$P_4(E) = 1 - 6E + 11E^2 - 6E^3 + E^4$$

are sufficient to define $P_n(E)$ for all $n > 2$. The first three terms of $P_n(E)$ are readily computed to be

$$P_n(E) = 1 - (2n - 2)\, E + (2n^2 - 7n + 7)\, E^2 + \ldots.$$

The formula to be developed yields the asymptotic expressions:

$$M(0) = e^{-2}\left(1 + \frac{1}{n} + \frac{1}{2n(n-1)}\right) + O(n^{-3})\,.$$

and

$$M(r) = \frac{e^{-2}2^r}{r!}\left(1 - \frac{(r^2 - r - 4)}{4n} + \frac{r^4 - 6r^3 + 3r^2 + 2r + 16}{32n(n-1)}\right) + O(n^{-3}),$$

A peculiarity arises here. There is a "pseudo recurrence formula"

$$P_n(E) = (1 - 2E)\,P_{n-1}(E) - E^2P_{n-2}(E),$$

which is *not* satisfied by the $P_n(E)$ associated with this example. Nevertheless, this "pseudo recurrence formula" yields the correct asymptotic series. The reason for this is that if correct values of $P_{n-1}(E)$ and $P_{n-2}(E)$ are substituted in the "pseudo recurrence formula" the formula yields the correct polynominal $P_n(E)$ except for the term in E^n. Asymptotically, this term is of no importance. The author has constructed several examples of problems which can be associated with "pseudo recurrence formulae" which are simpler than the true recurrence formulae and which yield the proper asymptotic series. However, no general theory of this phenomenon has as yet been formulated.

5. The general theory. From this point on, only permutation problems whose polynominal operators satisfy a difference equation of the type

$$\begin{aligned}(2)\quad P_n = {} & (\alpha_1 - \beta_1E)\,P_{n-1} + (\alpha_2 - \beta_2E + \gamma_2E^2)\,P_{n-2} + (\alpha_3 - \beta_3E \\ & + \gamma_3E^2 - \delta_3E^3)\,P_{n-3} \\ & + \ldots + (\alpha_k - \beta_kE + \gamma_kE^2 + \ldots + (-1)^k\lambda_kE^k)\,P_{n-k}(E).\end{aligned}$$

where $P_n = P_n(E)$, k is a fixed integer and all the Greek letters are constants, will be considered. It does not seem possible to give a precise characterization of the problems whose operators satisfy such a recurrence formula but some relevant observations may be made here. The total number of conditions possible is n^2. If in a specific problem the number of conditions in the set S is of the form $an^2 + bn + c$ with $a \neq 0$, no recurrence formula of the above type is possible. It is also true that there is no asymptotic series of the type

$$M(r) = \frac{e^{-A}A^r}{r!}\left(1 + \frac{c_1}{n} + \ldots\right).$$

Examples of such problems have been given by Kaplansky and Riordan in **(6)** but they are outside our scope. If the set S had $kn + l$ conditions a recurrence formula of the given type is possible and if these conditions form a reasonably regular pattern of marked squares in the chessboard representation the existence of a suitable recurrence is likely, and probably can be obtained in a routine way, by the methods given in **(6)** or **(7)**.

Assume now that a recurrence formula for $P_n(E)$ exists and is given by equation (2). An asymptotic series of the type given in equation (1) is sought. In this connection it is possible to show that the method given by Kaplansky in **(3)** yields a result of the form

$$M(r) = \frac{e^{-A}A^r}{r!}\left(1 + \frac{c}{n}\right) + O(n^{-2})$$

in those cases where $\alpha_i > 0$, $\beta_i > 0$ $(i = 1, 2, 3, \ldots, k)$. In what follows it is assumed that the complete asymptotic series exists and the work is confined to the computation of the c_i under this assumption. The polynomial $P_n(E)$ is given by:

$$P_n(E) = 1 - \phi_{n,1} E + \phi_{n,2} E^2 + \ldots + (-1)^n \phi_{n,n} E^n. \tag{3}$$

Using equations (2) and (3) it follows by complete induction that $\phi_{n,i}$ is a polynominal of degree i in n with coefficients which are functions of i. It is convenient to express $\phi_{n,i}$ in the form

$$\phi_{n,i} = C_0^{(i)} n_i + C_1^{(i)} (n-1)_{i-1} + C_2^{(i)} (n-2)_{i-2} + \ldots + C_i^{(i)} \tag{4}$$

The notation n_i is the Jordan factorial notation defined previously. To avoid complications of notation we define $\phi_{n,0} = 1$ and $\phi_{n,r} = 0$ if $r < 0$ or $r > n$. On substituting the expression for $P_n(E)$ as given by (3) into the recurrence (2) and arranging the result in powers of E the following relations are obtained:

$$\sum_{i=1}^{k} \alpha_i = 1,$$

(from the constant term), and

$$\phi_{n,r} = \sum_{i=1}^{k} \alpha_i \phi_{n-i,r} + \sum_{i=1}^{k} \beta_i \phi_{n-i,r-1} + \sum_{i=2}^{k} \gamma_i \phi_{n-i,r-2} + \ldots + \lambda_k \phi_{n-k,r-k}. \tag{5}$$

The expression (4) for $\phi_{n,i}$ is substituted into equation (5) to yield an expression which will be referred to as equation (6). Because of its extreme length, equation (6) is not written down here. On comparing coefficients of n^r in equation (6) and by the use of induction the following result is obtained:

$$C_0^{(r)} = \frac{A^r}{r!}, \quad A = \frac{\beta_1 + \beta_2 + \ldots + \beta_k}{\alpha_1 + 2\alpha_2 + \ldots + k\alpha_k}. \tag{7}$$

To compute $C_1^{(r)}$ the following procedure is used. First, all the $\phi_{n,i}$ occurring in equation (6) are expressed as factorial polynomials in terms of the variable $n - k - 2$ by making use of the relationship

$$(n+t)_u = n_u + tu\, n_{u-1} + \frac{t(t-1)\, u(u-1)}{2!} n_{u-2} - \ldots .$$

Then the coefficients of $(n - k - 2)_{r-2}$ in both sides of the resultant equation are equated to yield

$$C_1^{(r)} = \frac{A}{r-1} C_1^{(r-1)} - B \frac{A^{r-2}}{(r-1)!}, \tag{8}$$

where

$$B = \frac{1}{P}(KA^2 + LA + M),$$

$$K = \left\{ \sum_{i=1}^{k} \alpha_i \binom{k-i+2}{2} \right\} - \binom{k+2}{2},$$

$$L = \sum_{i=1}^{k} (k - i + 2)\,\beta_i,$$

$$M = \sum_{i=2}^{k} \gamma_i,$$

$$P = \left\{\sum_{i=1}^{k} (k - i + 1)\,\alpha_i\right\} - (k + 1).$$

The recurrence formula for $P_n(E)$ does not determine $C_1^{(1)}$ but this can be obtained from a knowledge of $P_1(E)$. In terms of $C_1^{(1)}$, B and A it is easily seen by induction that equation (8) has as solution

$$C_1^{(r)} = \frac{A^{r-1}C_1^{(1)}}{(r-1)!} - \frac{A^{r-2}B}{(r-2)!}. \tag{9}$$

At this point the first two terms of the asymptotic series will be computed. From the relationship $M(r) = P_n(E)\, g_r^*(0)$, the result

$$M(r) = \frac{1}{n!}\left\{\phi_{n,r}\binom{r}{r}(n-r)! - \phi_{n,r+1}\binom{r+1}{r}(n-r-1)! \right.$$
$$\left. + \phi_{n,r+2}\binom{r+2}{r}(n-r-2)! - \dots\right\}$$

is obtained. This reduces (on substituting for $\phi_{n,r}$ the expression (4)) to:

$$M(r) = \left\{C_0^{(r)} - \binom{r+1}{1}C_0^{(r+1)} + \binom{r+2}{2}C_0^{(r+2)} - \dots\right\}$$
$$- \frac{1}{n}\left\{C_1^{(r)} - \binom{r+1}{1}C_1^{(r+1)} + \binom{r+2}{2}C_1^{(r+2)} - \dots\right\}$$
$$+ O(n^{-2}).$$

Substituting for $C_0^{(i)}$ and $C_1^{(i)}$ the expression obtained in (7) and (8) yields the equation

$$M(r) = \left\{\frac{A^r}{r!} - \binom{r+1}{1}\frac{A^{r+1}}{(r+1)!} + \binom{r+2}{2}\frac{A^{r+2}}{(r+2)!} - \dots\right\}$$
$$+ \frac{1}{n}\left[C_1^{(1)}\left\{\frac{A^{r-1}}{(r-1)!} - \binom{r+1}{1}\frac{A^r}{r!} + \binom{r+2}{2}\frac{A^{r+1}}{(r+1)!} - \dots\right\}\right.$$
$$\left. - B\left\{\frac{A^{r-2}}{(r-2)!} - \binom{r+1}{1}\frac{A^{r-1}}{(r-1)!} + \binom{r+2}{2}\frac{A^r}{r!} - \dots\right\}\right]$$
$$+ O(n^{-2}).$$

This reduces to

$$M(r) = \frac{e^{-A}A^r}{r!}\left[1 + \frac{1}{n}\left\{C_1^{(1)}\left(\frac{r}{A} - 1\right) - B\left(\frac{r(r-1)}{A^2} - \frac{2r}{A} + 1\right)\right\}\right] + O(n^{-2}).$$

In the case where $r = 0$, the result further reduces to

$$M(0) = e^{-A}\left[1 - \frac{B + C_1^{(1)}}{n}\right] + O(n^{-2}).$$

The author has computed two further terms of the asymptotic series. The results are quite complicated in form. The term in $1/n(n-1)$ has been verified by means of the examples quoted previously. It is given here for completeness but all the computations have been omitted as they are quite involved and do not utilize any new idea. The final formula contains the number $C_2^{(2)}$ which is not obtainable from the recurrence formula for $P_n(E)$. All that is required for the computation of $C_2^{(2)}$ is a knowledge of $P_3(E)$. The final result is

$$M(0) = e^{-A}\left[1 - \frac{1}{n}(B + C_1^{(1)}) + \frac{1}{n(n-1)}\left\{C_2^{(2)} - \frac{\Gamma + B^2}{A} + \frac{B^2}{2}\right\}\right] + O(n^{-3}),$$

$$M(r) = \frac{e^{-A}A^r}{r!}\left[1 + \frac{1}{n}\left\{C_1^{(1)}\left(\frac{r}{A} - 1\right) - B\left(\frac{r(r-1)}{A^2} - \frac{2r}{A} + 1\right)\right\}\right.$$
$$+ \frac{1}{n(n-1)}\left\{C_2^{(2)}\left(\frac{r(r-1)}{A^2} - \frac{2r}{A} + 1\right)\right.$$
$$+ \frac{\Gamma}{A}\left(\frac{r(r-1)(r-2)}{A^3} - \frac{3r(r-1)}{A^2} + \frac{3r}{A} - 1\right)$$
$$\left.\left. + B^2\left(\frac{r(r-1)^2(r-2)}{2A^4} - \frac{r(r-1)(2r-1)}{A^3} + \frac{3r^2}{A^2} - \frac{2r+1}{A} + \frac{1}{2}\right)\right\}\right]$$
$$+ O(n^{-3}).$$

All terms in $M(r)$ except Γ have been previously defined. The value of Γ is given by the expression:

$$\Gamma = -\frac{1}{P}\{(RA^4 + SA^3 + TA^2 + UA) + C_1^{(1)}(KA^3 + LA^2 + MA) - B(KA^2 - M)\},$$

where

$$R = \left\{\sum_{i=1}^{k} \alpha_i \binom{k-i+3}{3}\right\} - \binom{k+3}{3},$$

$$S = \sum_{i=1}^{k} \beta_i \binom{k-i+3}{2},$$

$$T = \sum_{i=2}^{k} \gamma_i (k - i + 3),$$

$$U = \sum_{i=3}^{k} \delta_i.$$

The above formulae while formidable in appearance are quite simple to apply in practical cases. In none of the five examples quoted did the computations require as much as five minutes.

6. Distribution moments. The difference equation for $P_n(E)$ may be used to yield all the moments of the distribution of $M(r)$ as well as the asymptotic series. In this section formulae are established for the mean m and the variance v of the distribution.

The computation of moments is most easily carried out by the use of the notion of a factorial moment. The factorial moments are more natural to the type of problem considered in this paper than are the more usual power moments. The ith factorial moment of the distribution of the number of hits is defined as $M^{(i)}$, where

$$M^{(i)} = \sum_{r=0}^{n} r(r-1)(r-2)\ldots(r-i+1)\,M(r).$$

It has been shown in **(4)** and **(3)** that

$$M^{(i)} = \phi_{n,i}\Big/\binom{n}{i}.$$

Actually, this result follows easily by a direct computation. In terms of these factorial moments the mean m and the variance v are given by:

$$m = M^{(1)}, \quad v = M^{(2)} + M^{(1)} - \{M^{(1)}\}^2.$$

In terms of the constants computed in this paper these formulae become:

$$m = \frac{\phi_{n,1}}{n} = C_0^{(1)} + \frac{C_1^{(1)}}{n} = A + \frac{C_1^{(1)}}{n},$$

$$\begin{aligned}
v &= M^{(2)} + M^{(1)} - \{M^{(1)}\}^2 \\
&= \frac{2\phi_{n,2}}{n(n-1)} + \left(A + \frac{C_1^{(1)}}{n}\right) - \left(A + \frac{C_1^{(1)}}{n}\right)^2 \\
&= 2C_0^{(2)} + \frac{2C_1^{(2)}}{n} + \frac{2C_2^{(2)}}{n(n-1)} + A + \frac{C_1^{(1)}}{n} - \left(A + \frac{C_1^{(1)}}{n}\right)^2 \\
&= A^2 + \frac{2}{n}(AC_1^{(1)} - B) + \frac{2C_2^{(2)}}{n(n-1)} + A + \frac{C_1^{(1)}}{n} - \left(A + \frac{C_1^{(1)}}{n}\right)^2 \\
&= A + \frac{1}{n}(C_1^{(1)} - 2B) - \frac{\{C_1^{(1)}\}^2}{n^2} + \frac{2C_2^{(2)}}{n(n-1)}.
\end{aligned}$$

References

1. T. S. Broderick, *On some symbolic formulae in probability theory*, Proc. Roy. Irish Acad., *44* (1937), 19–28.
2. M. Fréchet, *Les probabilités associées à un système d'événements compatibles et dépendents*, Actualités Scientifiques et Industrielles, nos 859 et 942 (Paris, 1940 and 1943).
3. I. Kaplansky, *Symbolic solution of certain problems in permutations*, Bull. Amer. Math. Soc., *50* (1944), 906–914.
4. ———, *The asymptotic distribution of runs of consecutive elements*, Ann. Math. Statist., *16* (1945), 200–203.
5. I. Kaplansky and J. Riordan, *Le problème des ménages*, Scripta Math., *12* (1946), 113–124.
6. ———, *The problem of the rooks and its applications*, Duke Math. J., *16* (1946), 259–268.
7. N. S. Mendelsohn, *Symbolic solution of card matching problems*, Bull. Amer. Math. Soc., *52* (1946), 918–924.
8. ———, *Applications of combinational formulae to generalisations of Wilson's theorem*, Can. J. Math., *1* (1949), 328–336.
9. J. Riordan, *Three line Latin rectangles* II, Amer. Math. Monthly, *53* (1946), 18–20.

University of Manitoba

MAXIMAL DETERMINANTS IN COMBINATORIAL INVESTIGATIONS

H. J. RYSER

1. Introduction. Let Q be a matrix of order v, all of whose entries are 0's and 1's. Let the total number of 1's in Q be t, and let the absolute value of the determinant of Q be denoted by $|\det Q|$. In this paper we study the problem of determining the maximum of $|\det Q|$ for fixed t and v. It turns out that this problem is closely related to the v, k, λ problem, which has been extensively studied of late.

A v, k, λ configuration is defined as an arrangement of v elements $x_1, x_2, \ldots, x_v$ into v sets $S_1, S_2, \ldots, S_v$ such that each set contains exactly k distinct elements and such that each pair of sets has exactly λ elements in common $(0 < \lambda < k < v)$. If element x_j belongs to set S_i, let $a_{ij} = 1$; and if x_j does not belong to S_i, let $a_{ij} = 0$. The v by v matrix $A = [a_{ij}]$ is called the *incidence matrix* of the v, k, λ configuration. These matrices have been very useful in establishing the nonexistence of certain configurations (**1**; **2**). A general survey of the literature pertaining to v, k, λ configurations may be found in (**4**). In particular one proves that in a v, k, λ configuration,

$$k - \lambda = k^2 - \lambda v$$

and

$$AA^T = A^TA = B.$$

Here A^T denotes the transpose of the incidence matrix A, and the matrix B has k in the main diagonal and λ in all other positions. It is easy to see that $\det B = k^2(k-\lambda)^{v-1}$, whence it follows that

$$|\det A| = k(k-\lambda)^{\frac{1}{2}(v-1)}.$$

2. Theorems on maximal determinants.

THEOREM 1. *Let Q be a* 0, 1 *matrix of order v, containing exactly t* 1*'s. Let k denote a positive real, and set $\lambda = k(k-1)/(v-1)$. If $t < kv$ and $0 < \lambda < k - \lambda$, or if $t > kv$ and $0 < k - \lambda < \lambda$, then*

$$|\det Q| < k(k-\lambda)^{\frac{1}{2}(v-1)}.$$

Let E be a 0, 1 matrix. Let $E(x, y)$ denote the matrix formed from E by replacing each 1 of E by x and each 0 of E by y, where x and y are indeterminates. Using this notation, we may write

$$Q_1 = Q(-(k-\lambda)/\lambda, 1).$$

Received May 31, 1955.

Now set $p = (k - \lambda)/\lambda$, and define the matrix $\bar{Q}$ of order $v + 1$ by

$$\bar{Q} = \begin{bmatrix} p & z \\ z^T & Q_1 \end{bmatrix}, \tag{1}$$

where $z = (\sqrt{p}, \ldots, \sqrt{p})$. By the Hadamard determinant theorem,

$$|\det \bar{Q}| \leq \sqrt{p^2 + vp} \prod_{i=1}^{v} \sqrt{p + s_i}\,, \tag{2}$$

where s_i denotes the sum of the squares of the ith row of Q_1. Now

$$p^2 + vp = p\left(\frac{k - \lambda + \lambda v}{\lambda}\right) = \frac{k^2}{\lambda^2}(k - \lambda).$$

Moreover,

$$s_1 + \ldots + s_v = tp^2 + (v^2 - t) = t(p^2 - 1) + v^2.$$

By hypothesis, $t \leq kv$ and $p^2 \geq 1$, or $t \geq kv$ and $p^2 \leq 1$. Hence we may conclude

$$s_1 + \ldots + s_v \leq kv(p^2 - 1) + v^2.$$

Now introduce quantities $\bar{s}_i$ such that

$$\bar{s}_i \geq s_i$$

and

$$\bar{s}_1 + \ldots + \bar{s}_v = v(kp^2 + v - k). \tag{3}$$

By (3),

$$\begin{aligned} \sum_{i=1}^{v} (p + \bar{s}_i) &= v(kp^2 + v - k + p) = v[kp^2 + (\lambda v - \lambda k + k - \lambda)/\lambda] \\ &= vkp(p + 1) = v(k - \lambda)k^2/\lambda^2. \end{aligned}$$

Since the geometric mean of v positive quantities is less than or equal to their arithmetic mean, we may write

$$\prod_{i=1}^{v} (p + \bar{s}_i) \leq \left(\frac{1}{v}\sum_{i=1}^{v} (p + \bar{s}_i)\right)^v, \tag{4}$$

whence

$$\prod_{i=1}^{v} (p + \bar{s}_i) \leq (k - \lambda)^v k^{2v}/\lambda^{2v}. \tag{5}$$

Hence by (2),

$$\begin{aligned} |\det \bar{Q}| &\leq \frac{k}{\lambda}\sqrt{k - \lambda} \prod_{i=1}^{v} \sqrt{p + \bar{s}_i} \\ &\leq \frac{k}{\lambda}\sqrt{k - \lambda}\left(\frac{k}{\lambda}\sqrt{k - \lambda}\right)^v = \left(\frac{k}{\lambda}\sqrt{k - \lambda}\right)^{v+1}. \end{aligned} \tag{6}$$

To evaluate det $\bar{Q}$, multiply row one by $-1/\sqrt{p}$ and add the resulting row to each of the other rows. From (6) it follows that

$$|\det \bar{Q}| = p|\det Q(-k/\lambda, 0)| < (k\sqrt{k-\lambda}/\lambda)^{v+1}. \tag{7}$$

But

$$|\det Q(-k/\lambda, 0)| = (k/\lambda)^{v}|\det Q|, \text{ whence}$$

$$p|\det Q| < \frac{k}{\lambda}(\sqrt{k-\lambda})^{v+1},$$

and

$$|\det Q| < k(\sqrt{k-\lambda})^{v-1}.$$

Using the notation of Theorem 1, we have

THEOREM 2. *If* $|\det Q| = k(k-\lambda)^{\frac{1}{2}(v-1)}$, *then* Q *is the incidence matrix of a* v, k, λ *configuration.*

If equality holds in Theorem 1, then

$$p\left|\det Q\left(-\frac{k}{\lambda}, 0\right)\right| = \left(\frac{k\sqrt{k-\lambda}}{\lambda}\right)^{v+1},$$

and by (7),

$$|\det \bar{Q}| = (k\sqrt{k-\lambda}/\lambda)^{v+1}. \tag{8}$$

Equality in (6) implies equality in (5) and (4). But for equality to hold in (4), we must have

$$p + \bar{s}_i = (k-\lambda)k^2/\lambda^2.$$

But then the equality in (6) implies

$$\bar{Q}\bar{Q}^{T} = \frac{k^2(k-\lambda)}{\lambda^2} I, \tag{9}$$

where I is the identity matrix of order $v+1$. Thus

$$Q_1Q_1^{T} = \frac{k^2}{\lambda^2}(k-\lambda)\, I - pS, \tag{10}$$

where $Q_1 = Q(-p, 1)$, and S is the v by v matrix of all 1's. Let e denote the number of 1's in row r of Q. Then

$$p^2e + (v-e)\cdot 1 = \frac{k^2}{\lambda^2}(k-\lambda) - p,$$

and

$$(p^2-1)\,e = \frac{k^2}{\lambda^2}(k-\lambda) - p - v,$$

whence we conclude that $e = k$. Let f denote the inner product of rows r and s of Q, where $r \neq s$. Then

$$fp^2 - 2(k-f)p + v - 2k + f = -p,$$

whence

$$f(p^2 + 2p + 1) = 2kp - p + 2k - v,$$

and $fk^2/\lambda^2 = k^2/\lambda$. Thus $f = \lambda$, and Q is the incidence matrix of a v, k, λ configuration.

It is now clear that we have established the following:

THEOREM 3. *Let Q be a* 0, 1 *matrix of order v, containing exactly t 1's. Let $k = t/v$ and set $\lambda = k(k-1)/(v-1)$, with $0 < \lambda < k < v$. Then*

$$|\det Q| \leq k(k-\lambda)^{\frac{1}{2}(v-1)},$$

and equality holds if and only if Q is the incidence matrix of a v, k, λ configuration.

Consider once again Theorem 1. Note that $(k-\lambda)/\lambda = (v-k)/(k-1)$. Thus the requirement $\lambda \leq k - \lambda$ means $k \leq \frac{1}{2}(v+1)$, and $k - \lambda \leq \lambda$ means $k \geq \frac{1}{2}(v+1)$. Suppose that $k = \frac{1}{2}(v+1)$. Then if Q is a 0, 1 matrix with no restriction on the number of 1's, we must have

$$|\det Q| \leq \frac{(v+1)^{\frac{1}{2}(v+1)}}{2^v}. \tag{11}$$

The incidence matrix associated with the case of equality has parameters $v = 4\lambda - 1$, $k = 2\lambda$, $\lambda = \lambda$. These incidence matrices give rise to the Hadamard matrices of order 4λ **(3)**. The determination of the maximum of $|\det Q|$, where Q is of arbitrary order v, is an unsolved problem of considerable difficulty **(5)**.

If we place no restriction on the number of 1's in the 0, 1 matrix Q of order v and assume that $|\det Q| = k(k-\lambda)^{\frac{1}{2}(v-1)}$, then we may not conclude in general that Q is the incidence matrix of a v, k, λ configuration. For example, let A be an incidence matrix of a v, k, λ configuration with $v - 2k > 0$. Define its complement C by $A + C = S$, where S is the matrix of all 1's. The complement of A is again a v, k, λ configuration with parameters $\bar{v} = v$, $\bar{k} = v - k$, and $\bar{\lambda} = v - 2k + \lambda$. Note that

$$|\det C| = (v-k)(k-\lambda)^{\frac{1}{2}(v-1)}.$$

It is easy to check that

$$A^{-1} = \frac{1}{(k-\lambda)}\left(A^T - \frac{\lambda}{k} S\right),$$

where A^{-1} denotes the inverse of A. Thus in $A = [a_{rs}]$, if $a_{rs} = 1$, then the cofactor of a_{rs},

$$A_{rs} = \frac{1}{k} \det A.$$

Similarly for the complement $C = [c_{rs}]$, if $c_{rs} = 1$, then the cofactor of c_{rs},

$$C_{rs} = \frac{1}{v-k}\det C.$$

We are assuming that $v - 2k > 0$. Thus we may replace $v - 2k$ of the 1's in the first row of C by 0's. The resulting matrix Q is a 0, 1 matrix satisfying

$$|\det Q| = k(k-\lambda)^{\frac{1}{2}(v-1)},$$

but Q is not an incidence matrix of a v, k, λ configuration.

References

1. R. H. Bruck and H. J. Ryser, *The nonexistence of certain finite projective planes*, Can. J. Math., *1* (1949), 88–93.
2. S. Chowla and H. J. Ryser, *Combinatorial problems*, Can. J. Math., *2* (1950), 93–99.
3. R. E. A. C. Paley, *On orthogonal matrices*, J. Math. Phys., *12* (1933), 311–320.
4. H. J. Ryser, *Geometries and incidence matrices*, Slaught Memorial Papers (Suppl. Amer. Math. Monthly), *62* (1955), 25–31.
5. John Williamson, *Determinants whose elements are* 0 *and* 1, Amer. Math. Monthly, *53* (1946), 427–434.

Ohio State University

ON A CLASS OF ALMOST ALTERNATIVE ALGEBRAS

L. A. KOKORIS

Introduction. In the study of almost alternative algebras (**2**) relative to quasiequivalence an important class called algebras of (γ, δ) type arises. An algebra of (γ, δ) type is a finite dimensional algebra $\mathfrak{A}$ over a field $\mathfrak{F}$ satisfying the identities

$$(1)\qquad z(xy) = (zx)y + \gamma(xz)y - \gamma x(zy) + \delta(yz)x - \delta y(zx),$$

and

$$(2)\qquad (xy)z = x(yz) + \gamma(xz)y - \gamma x(zy) + (\delta - 1)(yz)x - (\delta - 1)y(zx)$$

where γ and δ are elements of $\mathfrak{F}$ satisfying $\gamma^2 - \delta^2 + \delta = 1$. We shall restrict our study to (γ, δ) type algebras with characteristic $\neq 2$, 3, or 5 and with $\delta \neq 0, 1$. With these restrictions the algebras are power-associative. Also, Albert has shown (**2**, p. 36) that if an algebra $\mathfrak{A}$ of (γ, δ) type has an idempotent e it can be decomposed into a supplementary sum $\mathfrak{A} = \mathfrak{A}_{11} + \mathfrak{A}_{10} + \mathfrak{A}_{01} + \mathfrak{A}_{00}$ where x is in $\mathfrak{A}_{ij}$ if and only if $ex = ix$ and $xe = jx$. The subspaces of our decomposition have the same multiplicative properties as in the case of an associative algebra.

The concepts of a solvable algebra, nilpotent algebra, and nil algebra are equivalent for (γ, δ) type algebras with the restrictions mentioned above (**2**, p. 35). The radical is defined to be the maximal nilideal and it is then proved that a simple algebra is either associative or contains a unity which is an absolutely primitive idempotent. A semisimple algebra is a direct sum of simple algebras.

If $\delta = 0$ or 1 we have the four pairs $(\gamma, \delta) = (1, 1)$, $(-1, 0)$, $(1, 0)$, or $(-1, 1)$. The pair $(-1, 1)$ implies that the algebra is right alternative and $(1, 0)$ implies the left alternative law. In the remaining two cases we are not able to obtain the same multiplicative relations for the subspaces of the decomposition as for the general case and it seems that the results here should be different.

1. Decomposition relative to an idempotent. Let $\mathfrak{A}$ be an algebra of (γ, δ) type with characteristic $\neq 2$ and with an idempotent e. If $(\gamma, \delta) \neq (-1, 1)$ or $(1, 0)$, it is known that $\mathfrak{A}$ may be decomposed into a vector space direct sum $\mathfrak{A} = \mathfrak{A}_{11} + \mathfrak{A}_{10} + \mathfrak{A}_{01} + \mathfrak{A}_{00}$. This is the decomposition of the theory of associative algebras and we are able to obtain the multiplicative relations of the associative theory when $\delta \neq 0, 1$.

Received July 25, 1955. Presented to the American Mathematical Society, September 1, 1955.

THEOREM 1. *Let $\mathfrak{A}$ be an algebra of (γ, δ) type with $\delta \neq 0, 1$ and characteristic $\neq 2, 3$. Then $\mathfrak{A}_{ij}\mathfrak{A}_{qt} = 0$ if $j \neq q$ and $\mathfrak{A}_{ij}\mathfrak{A}_{jt} < \mathfrak{A}_{it}$.*

The proof is made by considering the various cases. Take $z = e$, x in $\mathfrak{A}_{ij}$, and y in $\mathfrak{A}_{qt}$. Then (1) becomes

$$e(xy) = (i + j\gamma - q\gamma)xy + (t\delta - i\delta)yx. \tag{3}$$

Interchanging x and y gives

$$e(yx) = (q + t\gamma - i\gamma)yx + (j\delta - q\delta)xy. \tag{4}$$

With x, y, z as above, (2) becomes

$$(xy)e = (t + j\gamma - q\gamma)xy + (\delta - 1)(t - i)yx. \tag{5}$$

Interchanging the roles of x and y we have

$$(yx)e = (j + t\gamma - i\gamma)yx + (\delta - 1)(j - q)xy. \tag{6}$$

Now consider the case where x and y are in $\mathfrak{A}_{11}$ so that $i = j = q = t = 1$. Relations (3) and (5) yield $e(xy) = xy$, and $(xy)e = xy$. Therefore, $\mathfrak{A}_{11}$ is a subalgebra. The values $i = j = q = t = 0$ in (3) and (5) prove that $\mathfrak{A}_{00}$ is also a subalgebra. When x is in $\mathfrak{A}_{11}$, y is in $\mathfrak{A}_{00}$, (3) and (4) give $e(xy) = (1+\gamma)xy - \delta yx$ and $e(yx) = -\gamma yx + \delta xy$. We now use the fact that $L_e^2 = L_e$ (later (cf. **2**, p. 36) we shall also need $R_e^2 = R_e$) to see that

$$e[e(xy)] = e(xy), \quad e(xy) = (1 + \gamma)[e(xy)] - \delta e(yx).$$

It follows that $(\gamma^2 - \delta^2 + \gamma)xy = 0$. Since $\gamma^2 - \delta^2 + \gamma = 0$ together with the defining relation $\gamma^2 - \delta^2 + \delta = 1$ for an algebra of (γ, δ) type implies $\delta = 0$, we must have $xy = 0$. Also,

$$e[e(yx)] = e(yx), \quad -\gamma e(yx) + \delta e(xy) = e(yx).$$

Consequently $(\gamma^2 - \delta^2 + \gamma)yx = 0$ and so $yx = 0$. Thus $\mathfrak{A}_{11}$ and $\mathfrak{A}_{00}$ are orthogonal subalgebras.

If x is in $\mathfrak{A}_{11}$ and y is in $\mathfrak{A}_{10}$, we have $e(xy) = xy - \delta yx$, $e(yx) = (1 - \gamma)yx = (yx)e$, and $(xy)e = (1 - \delta)yx$. Then $e[e(xy)] = e(xy)$ implies $\delta e(yx) = 0$ and it follows that $yx = 0$. This also proves that xy is in $\mathfrak{A}_{10}$. Next let x be in $\mathfrak{A}_{11}$ and y be in $\mathfrak{A}_{01}$ so that

$$e(xy) = (1 + \gamma)xy = (xy)e, \quad e(yx) = \delta xy, \quad (yx)e = yx + (\delta - 1)xy.$$

The result $xy = 0$ is obtained by noting that $[(yx)e]e = (yx)e$ and $(\delta - 1)[(xy)e] = 0$. Then yx is in $\mathfrak{A}_{01}$.

Consider the case where x and y are both in $\mathfrak{A}_{10}$ and

$$e(xy) = (1 - \gamma)xy - \delta yx, \quad e(yx) = (1 - \gamma)yx - \delta xy, \quad (xy)e = -\gamma xy + (1 - \delta)yx, \quad (yx)e = -\gamma yx + (1 - \delta)xy.$$

From $e[e(xy)] = e(xy)$ and $e[e(yx)] = e(yx)$ we obtain $(\gamma + \delta)[e(xy) + e(yx)] = 0$. Since $\gamma + \delta \neq 0$ by hypothesis,

$$e(xy) + e(yx) = 0 = (1 - \gamma - \delta)(xy + yx).$$

Again $1-\gamma-\delta\neq 0$ by hypothesis, so $xy+yx=0$. Thus

$$e(xy)=(1-\gamma+\delta)xy,\ (xy)e=(-1-\gamma+\delta)xy.$$

Moreover,

$$e[e(xy)]=e(xy)=(1-\gamma+\delta)e(xy),\ (1-\gamma+\delta)(-\gamma+\delta)xy=0.$$

When the characteristic is not 3, $\delta\neq 0,1$ implies $xy=0$. The case with x in $\mathfrak{A}_{10}$ and y in $\mathfrak{A}_{01}$ is proved immediately by substituting in (3) to (6). If both x and y are in $\mathfrak{A}_{01}$, $e(xy)=\gamma xy+\delta yx$ and $e(yx)=\gamma yx+\delta xy$. Therefore

$$e[e(xy)]=e(xy)=\gamma e(xy)+\delta e(yx),\ e[e(yx)]=e(yx)=\gamma e(yx)+\delta e(xy)$$

when added give $(-1+\gamma+\delta)[e(xy)+e(yx)]=0$. Hence $(\gamma+\delta)(xy+yx)=0$ and thus $xy=-yx$. We then have

$$e(xy)=(\gamma-\delta)xy,\quad e[e(xy)]=e(xy)=(\gamma-\delta)[e(xy)].$$

This implies $(\gamma-\delta-1)(\gamma-\delta)xy=0$, $xy=0$.

Take x in $\mathfrak{A}_{10}$ and y in $\mathfrak{A}_{00}$. Then $e(xy)=xy-\delta yx$, $(xy)e=(1-\delta)yx$, and $(yx)e=-\gamma yx$. We have $[(xy)e]e]=(1-\delta)[(yx)e]=(xy)e$ and $(1-\delta)(1+\gamma)yx=0$. Our hypothesis on δ implies $yx=0$ and it follows that xy is in $\mathfrak{A}_{10}$. The last case is with x in $\mathfrak{A}_{01}$ and y in $\mathfrak{A}_{00}$. Relations (3) to (6) become

$$e(xy)=\gamma xy,\ e(yx)=\delta xy,\ (xy)e=\gamma xy,\ (yx)e=yx+(\delta-1)xy.$$

Also $e[e(yx)]=e(yx)=\delta e(xy)$ and so $\delta(1-\gamma)xy=0$. Since $\delta(1-\gamma)\neq 0$, $xy=0$ and yx is in $\mathfrak{A}_{01}$. This completes the proof of Theorem 1.

2. Power-associativity. When $x=y=z$, relation (1) becomes $(1+\gamma+\delta)(xx^2-x^2x)=0$ and (2) yields $(2-\gamma-\delta)(xx^2-x^2x)=0$. Addition of the two expressions gives $xx^2=x^2x$ if the characteristic $\neq 3$. Assume that $\mathfrak{A}$ is an algebra of (γ,δ) type with characteristic $\neq 2,3$ and let $z=x^2$, $y=x$ in (1) and (2) to obtain

$$x^2x^2=(1+\gamma+\delta)x^3x-(\gamma+\delta)xx^3=(-1+\gamma+\delta)x^3x-(-2+\gamma+\delta)xx^3.$$

It follows that $2x^3x=2xx^3$ and $x^2x^2=x^3x=xx^3$. If also $\mathfrak{A}$ has characteristic $\neq 5$, it satisfies the hypotheses of the known (**1**, Lemma 4):

LEMMA 1. *Let $\mathfrak{A}$ be an algebra with characteristic $\neq 2,3,5$ and $x^\lambda x^\mu=x^{\lambda+\mu}$ for $\lambda+\mu<n$, $n>5$. Then*

$$x^{n-\alpha}x^\alpha=x^{n-1}x+\frac{\alpha-1}{2}[x^{n-1},x]\qquad(\alpha=1,\dots,n-1)\tag{7}$$

where $[x^{n-1},x]=x^{n-1}x-xx^{n-1}$. *Also,* $n[x^{n-1},x]=0$.

The Lemma will be used to show that an algebra of (γ,δ) type is power-associative if its characteristic $\neq 2$, 3, 5. Write x^α for x, x^β for y and $x^{n-\alpha-\beta}$ for z in (1) where α,β are positive integers such that $\alpha+\beta<n$ and assume that $x^\lambda x^\mu=x^{\lambda+\mu}$ for $\lambda+\mu<n$ to obtain

$$x^{n-\alpha-\beta}x^{\alpha+\beta}=x^{n-\beta}x^\beta+\gamma x^{n-\beta}x^\beta-\gamma x^\alpha x^{n-\alpha}+\delta x^{n-\alpha}x^\alpha-\delta x^\beta x^{n-\beta}.$$

By (7) we have after multiplying by 2,

$$(\alpha+\beta-1)[x^{n-1},x] = (1+\gamma)(\beta-1)[x^{n-1},x] - \gamma(n-\alpha-1)[x^{n-1},x] \\ +\delta(\alpha-1)[x^{n-1},x] - \delta(n-\beta-1)[x^{n-1},x].$$

Thus either $[x^{n-1},x]=0$ or

$$\alpha+\beta-1 = \beta-1+\gamma\beta-\gamma+\gamma\alpha+\gamma+\delta\alpha-\delta+\delta\beta+\delta.$$

If $[x^{n-1},x]=0$, (7) implies $\mathfrak{A}$ is power-associative. Otherwise $\alpha=(\gamma+\delta)(\alpha+\beta)$. Since α and β are any positive integers, restricted only by $\alpha+\beta<n$, interchange α and β to obtain $\beta=(\gamma+\delta)(\alpha+\beta)$. Adding,

$$\alpha+\beta = 2\,(\gamma+\delta)(\alpha+\beta)$$

and $\alpha=\beta=1$ implies $2(\gamma+\delta)=1$. But it is impossible for γ and δ to satisfy both this equation and $\gamma^2-\delta^2+\delta=1$.

THEOREM 2. *An algebra $\mathfrak{A}$ of (γ,δ) type whose characteristic $\neq 2, 3, 5$ is power-associative.*

3. Simple algebras. From this point on we shall consider algebras of (γ,δ) type with $\delta\neq 0, 1$ and with characteristic $\neq 2, 3, 5$ so that we may use the results of Theorems 1 and 2. We shall make use of the associator (x, y, z) which is defined by $(x,y,z)=(xy)z-x(yz)$. If $\mathfrak{A}$ is an algebra with an idempotent e we may prove the following result.

LEMMA 2. *The associator (x, y, z) is 0 if one of the elements x, y, z is in $\mathfrak{A}_{10}$ or $\mathfrak{A}_{01}$.*

First consider the possible ordered triples with x_{10} in $\mathfrak{A}_{10}$ on the left and y, z in the decomposition subspaces. It is clear that by linearity we need only consider elements in the subspaces of the decomposition. By Theorem 1 it is clear that the only triples with x_{10} on the left giving nonzero products are

$$x_{10},\ y_{01},\ z_{11};\ x_{10},\ y_{01},\ z_{10};\ x_{10},\ y_{00},\ z_{01};\ x_{10},\ y_{00},\ z_{00},$$

where the subscripts indicate the subspaces in which the elements lie. Let $x=x_{10}$, $y=y_{01}$, $z=z_{11}$ in (2) and use the fact that our decomposition is supplementary to obtain $x_{10}(y_{01}z_{11})=(x_{10}y_{01})y_{11}$. Similarly we prove the result for the second and third triples. For the last triple we use (1) with $z=x_{10}$, $x=y_{00}$, $y=z_{00}$ to get $x_{10}(y_{00}z_{00})=(x_{10}y_{00})z_{00}$.

Triples with y_{10} in the middle giving nonzero products are

$$x_{11},\ y_{10},\ z_{01};\quad x_{11},\ y_{10},\ z_{00};\quad x_{01},\ y_{10},\ z_{01};\quad x_{01},\ y_{10},\ z_{00}.$$

The result of the Theorem is proved by making the obvious substitutions in (1) for the first two of these triples and in (2) for the last two.

There are also four triples with z_{10} on the right giving non-zero products. These are

$$x_{11},\ y_{11},\ z_{10};\ x_{01},\ y_{11},\ z_{10};\ x_{10},\ y_{01},\ z_{10};\ x_{00},\ y_{01},\ z_{10}.$$

For the first three substitute in (2) and use (1) for the last triple. By symmetry we have the result for elements in $\mathfrak{A}_{01}$.

COROLLARY. *The algebra $\mathfrak{A}$ is associative if and only if $\mathfrak{A}_{11}$ and $\mathfrak{A}_{00}$ are associative.*

Now let $\mathfrak{A}$ be a simple algebra. There must be a nonnilpotent element x in $\mathfrak{A}$ and the subalgebra generated by x must be associative since $\mathfrak{A}$ is power-associative. Since an associative algebra not a nilalgebra has an idempotent, $\mathfrak{A}$ has an idempotent e. Decompose $\mathfrak{A}$ relative to e. Then the sets

$$\mathfrak{B} = \mathfrak{A}_{11} + \mathfrak{A}_{10} + \mathfrak{A}_{01} + \mathfrak{A}_{01}\mathfrak{A}_{10}, \quad \mathfrak{C} = \mathfrak{A}_{00} + \mathfrak{A}_{10} + \mathfrak{A}_{01} + \mathfrak{A}_{10}\mathfrak{A}_{01}$$

can easily be seen to be ideals of $\mathfrak{A}$. Since e is in $\mathfrak{B}$, $\mathfrak{B} = \mathfrak{A}$ and thus $\mathfrak{A}_{00} = \mathfrak{A}_{01}\mathfrak{A}_{10}$. It follows from this and Lemma 2 that $\mathfrak{A}_{00}$ is zero or an associative algebra. In the latter case $\mathfrak{A}_{00}$ is simple for if $\mathfrak{B}_{00}$ were a proper ideal of $\mathfrak{A}_{00}$, then $\mathfrak{B}_{00}$ would generate the proper ideal

$$\mathfrak{B}_{00} + \mathfrak{A}_{10}\mathfrak{B}_{00} + \mathfrak{B}_{00}\mathfrak{A}_{01} + \mathfrak{A}_{10}\mathfrak{B}_{00}\mathfrak{A}_{01}$$

of $\mathfrak{A}$. The ideal $\mathfrak{C} = \mathfrak{A}$ or 0. If $\mathfrak{C} = \mathfrak{A}$, $\mathfrak{A}_{11} = \mathfrak{A}_{10}\mathfrak{A}_{01}$ and $\mathfrak{A}_{11}$ is a simple associative algebra. If $\mathfrak{C} = 0$, $\mathfrak{A} = \mathfrak{A}_{11}$ and e is the unity element of $\mathfrak{A}$. In case $e = u + v$ is not primitive, we can get a proper decomposition with respect to u and with the new $\mathfrak{A}_{00} \neq 0$. Then $\mathfrak{A}$ is associative. When e is not absolutely primitive we can find a scalar extension $\mathfrak{K}$ of the base field $\mathfrak{F}$ such that $e = u + v$ for pairwise orthogonal idempotents u, v in $\mathfrak{A}_{\mathfrak{K}}$. Consequently $\mathfrak{A}_{\mathfrak{K}}$ is associative and $\mathfrak{A}$ is associative.

THEOREM 3. *Let $\mathfrak{A}$ be a simple algebra of (γ, δ) type with $\delta \neq 0, 1$ and with characteristic $\neq 2, 3, 5$. Then $\mathfrak{A}$ is an associative algebra or $\mathfrak{A}$ has a unity quantity which is an absolutely primitive idempotent.*

4. Semisimple algebras.

The study of semisimple algebras begins with

THEOREM 4. *Let e be a principal idempotent of an algebra $\mathfrak{A}$ of (γ, δ) type with $\delta \neq 0, 1$ and characteristic $\neq 2, 3, 5$. Then $\mathfrak{A}_{10} + \mathfrak{A}_{01} + \mathfrak{A}_{00}$ is contained in the radical $\mathfrak{N}$ of $\mathfrak{A}$.*

The proof is made by an induction on the order of $\mathfrak{A}$. The result is clear when $\mathfrak{A}$ has order one. Assume the Theorem for all algebras of order less than n and let $\mathfrak{A}$ have order n. If $\mathfrak{A}$ is not semisimple we consider $\mathfrak{B} = \mathfrak{A} - \mathfrak{N}$ which has order $m < n$. The principal idempotent e of $\mathfrak{A}$ corresponds to a principal idempotent u of $\mathfrak{B}$. Decompose $\mathfrak{B}$ relative to u. Since $\mathfrak{B}$ is semisimple our induction hypothesis simplies $\mathfrak{B}_{10} + \mathfrak{B}_{01} + \mathfrak{B}_{00} = 0$. This implies that in the decomposition of $\mathfrak{A}$ relative to e, $\mathfrak{A}_{10} + \mathfrak{A}_{01} + \mathfrak{A}_{00} \subseteq \mathfrak{N}$.

If $\mathfrak{A}$ is simple, Theorem 3 implies $\mathfrak{A}$ has a unity e and an algebra with a unity has no other principal idempotent. Thus we may pass to the consideration of a semisimple algebra $\mathfrak{A}$ with a proper ideal $\mathfrak{B}$. The ideal $\mathfrak{B}$ can not be a nilideal so it must contain an idempotent and hence a principal idempotent e. Then $\mathfrak{B} = \mathfrak{B}_{11} + \mathfrak{B}_{10} + \mathfrak{B}_{01} + \mathfrak{B}_{00}$ and we may also decompose $\mathfrak{A}$ relative to e so that $\mathfrak{A} = \mathfrak{A}_{11} + \mathfrak{A}_{10} + \mathfrak{A}_{01} + \mathfrak{A}_{00}$. The idempotent e is in $\mathfrak{B}$ and so if

$ex = x$ or $xe = x$, it follows that x is also in $\mathfrak{B}$. Consequently, $\mathfrak{A} = \mathfrak{B}_{11} + \mathfrak{B}_{10} + \mathfrak{B}_{01} + \mathfrak{A}_{00}$. By the induction $\mathfrak{B}$ has radical $\mathfrak{M} = \mathfrak{M}_{11} + \mathfrak{B}_{10} + \mathfrak{B}_{01} + \mathfrak{B}_{00}$ where $\mathfrak{M}_{11}$ is the part of $\mathfrak{M}$ in $\mathfrak{B}_{11}$. Since $\mathfrak{B}$ is an ideal of $\mathfrak{A}$ it follows that $\mathfrak{M}$ is a nilideal of $\mathfrak{A}$ and that $\mathfrak{M} = 0$. Therefore $\mathfrak{A} = \mathfrak{B} \oplus \mathfrak{A}_{00}$ and e is the unity quantity of $\mathfrak{B}$. The subalgebra $\mathfrak{A}_{00}$ is an ideal of $\mathfrak{A}$ and by a repetition of the above argument $\mathfrak{A}_{00}$ has a unity f. Then $u = e + f$ is a unity for $\mathfrak{A}$ and is therefore the only principal idempotent of $\mathfrak{A}$. This completes the proof[1] of Theorem 4. We have also proved

THEOREM 5. *A semisimple algebra of* (γ, δ) *type with* $\delta \neq 0, 1$ *and with characteristic* $\neq 2, 3, 5$ *has a unity quantity and is a direct sum of simple algebras.*

[1]The reader should notice that our proofs follow those of Theorems 7 and 8 of Albert (3).

REFERENCES

1. A. A. Albert, *On the power-associativity of rings*, Summa Braziliensis Mathematicae, *2* (1948), 1–13.
2. ———, *Almost alternative algebras*, Portugaliae Mathematica, *8* (1949), 23–36.
3. ———, *A theory of power-associative commutative algebras*, Trans. Amer. Math. Soc., *69* (1950), 503–527.

Washington University, St. Louis.

ORTHOGONAL ISOMORPHIC REPRESENTATIONS OF FREE GROUPS

J. DE GROOT

1. Introduction. We consider the group $\mathfrak{G}$ of proper orthogonal transformations (rotations) in three-dimensional Euclidean space, represented by real orthogonal matrices (a_{ik}) $(i, k = 1, 2, 3)$ with determinant $+1$. It is known that this rotation group $\mathfrak{G}$ contains free (non-abelian) subgroups; in fact Hausdorff **(5)** showed how to find two rotations P and Q generating a group with only two non-trivial relations

$$P^2 = Q^3 = I.$$

Now the elements $PQPQ$ and PQ^2PQ^2 are free generators of a free rotation group R **(7)**. It was shown in **(4)**, starting from R and using transfinite induction, that $\mathfrak{G}$ contains even a free subgroup with continuously many free generators.[1]

Now it is clear that this method for constructing free subgroups of $\mathfrak{G}$ is an indirect one and furnishes only special free rotation groups. These disadvantages became more visible in certain problems (partly geometrical, partly group-theoretical) dealing with free rotation groups in spaces of dimension >3. Therefore we shall develop in this paper a straightforward and simple method of determining free subgroups of $\mathfrak{G}$ (we restrict ourselves, however, to the three-dimensional case). The only, but in many cases serious, difficulty with this type of problem is to prove time and again that certain products of matrices do not vanish identically. As our main result (Theorem II) we shall give, explicitly, continuously many rotations (with the same rotation angle, and rotation axes situated in the same plane), which are free generators of a free group (of continuous rank). Other representations of free groups were given by Fuchs-Rabinowitsch **(3)** and Doniakhi **(2**; see also Sanov **8)**, who use two-rowed square matrices; however, these cannot be orthogonal. Some conjectures are stated in §5.

Received March 23, 1955.

[1]After completion of this paper, the author heard from Poland that Sierpiński proved a lemma **(9**, 238**)** which, though not stated in terms of group theory, implies the existence of a free rotation group of countable rank, and from which the existence of a free rotation group of continuous rank can easily be deduced. Already, Sierpiński uses in his proof the "von Neumann numbers" (see §4 of this paper). On the other hand we see that Sierpiński's proof essentially makes use of the Hausdorff result **(5)**, just as does the proof in **(4)**. For this reason the methods derived in the present paper improve those in **(9)** and **(4)** and Theorems I and II cannot be obtained by employing the methods of **(9)**, **(4)** only.

2. Preliminaries. We consider polynomials V in the variables $\sin n\phi$, $\cos m\phi$ (n, m ranging over the integers) over the real field. Each term

$$\prod_i \sin^{r_i} n_i\phi \cdot \prod_j \cos^{s_j} m_j\phi$$

has degree

$$\sum_i |r_i n_i| + \sum_j |s_j m_j|.$$

The *degree* of V is the maximum of the degrees of each of its terms.

LEMMA I. *A polynomial V, having only one term of degree equal to the degree of V itself, is a non-constant function of ϕ.*

Proof. Using the formulae

$$\begin{aligned}\sin n\,\phi &= \binom{n}{1}\cos^{n-1}\phi\sin\phi - \binom{n}{3}\cos^{n-3}\phi\sin^3\phi + \dots\\ &= \left[\binom{n}{1} + \binom{n}{3} + \dots\right]\cos^{n-1}\phi\sin\phi + \dots,\\ \cos m\,\phi &= \cos^m\phi - \binom{m}{2}\cos^{m-2}\phi\sin^2\phi + \dots\\ &= \left[1 + \binom{m}{2} + \binom{m}{4} + \dots\right]\cos^m\phi + \dots,\end{aligned}$$

V is transformed into a polynomial in $\sin\phi$ and $\cos\phi$, again having only one term of maximal degree. This expression of V can obviously be written in the form

$$\sum_{i=0}^{k} (\alpha_i \sin\phi + \beta_i \cos\phi)\cos^{k-i}\phi + \gamma, \tag{1}$$

either α_0 or β_0 being equal to zero.

If $k = 0$ this polynomial does not vanish identically in ϕ. Then the lemma follows by induction. Suppose the lemma holds for $k - 1$; suppose that for the value k the polynomial $V \equiv 0$. Substituting $\phi = \pm\frac{1}{2}\pi$ in (1) we find $\alpha_k = \gamma = 0$; thus we can divide V by $\cos\phi$, and get a contradiction. So $V \not\equiv 0$, and V is non-constant.

LEMMA II. *Consider the real orthogonal matrices*

$$A(\phi) = \begin{pmatrix}\cos\phi & -\sin\phi & 0\\ \sin\phi & \cos\phi & 0\\ 0 & 0 & 1\end{pmatrix}, \quad B(\phi) = \begin{pmatrix}1 & 0 & 0\\ 0 & \cos\phi & -\sin\phi\\ 0 & \sin\phi & \cos\phi\end{pmatrix}.$$

Any proper[2] *product in terms of A and B is a non-constant matrix (depending on ϕ).*

[2]The product is proper if it cannot be transformed into the unity-matrix I using only the trivial relations $A\cdot A^{-1} = B\cdot B^{-1} = I$, $A\cdot I = I\cdot A = A$, $B\cdot I = I\cdot B = B$.

Proof. At first we prove the Lemma for products of the form

$$A^{n_1} B^{m_1} A^{n_2} B^{m_2} \dots A^{n_k} B^{m_k} \qquad (n_i, m_i \text{ integers} \neq 0) \tag{2}$$

If $k = 1$, we get

$$A^{n_1} B^{m_1} = \begin{pmatrix} \cos n_1\phi & -\sin n_1\phi \cos m_1\phi & \sin n_1\phi \sin m_1\phi \\ \sin n_1\phi & \cos n_1\phi \cos m_1\phi & -\cos n_1\phi \sin m_1\phi \\ 0 & \sin m_1\phi & \cos m_1\phi \end{pmatrix}.$$

This is obviously a non-constant matrix. (2) is a matrix (a_{ij}). Denote the degree of a_{ij} by d_{ij}. Suppose these degrees satisfy for $k = l - 1$ the relations

$$\begin{cases} d_{11}, d_{21} < \sum_{i=1}^{k} |n_i| + \sum_{i=1}^{k-1} |m_i|, \\ d_{12}, d_{13}, d_{22}, d_{23} = \sum_{i=1}^{k} (|n_i| + |m_i|), \end{cases} \tag{3}$$

while, moreover, each of the elements $a_{12}, a_{13}, a_{22}, a_{23}$ has exactly one term of the corresponding degree denoted in (3).

Now one sees easily by multiplying (a_{ij}) (for $k = l - 1$) with

$$A^{n_k} B^{m_k}$$

that (a_{ij}) (for $k = l$) satisfies the same conditions. Since this is also true for $k = 1$, the Lemma follows by induction, applying Lemma I, for all products of type (2).

If we multiply (2) on the right by

$$A^{n_{k+1}}$$

we see also in the same way—using the properties mentioned—that this product depends on ϕ. Since the interchanging of A and B in (2) does no harm, the Lemma is proved.

Remark. In the proof it is possible, but not necessary, to consider the whole of the matrix (d_{ik}); one could also deal with the degrees of the second row only. One might also consider the degree of the trace of (a_{ik}) (independence of the chosen coordinate-system).

3. Countable representations. Using the well-known substitution

$$\phi = 2 \arctan x \qquad (0 < \phi < \pi), \tag{4}$$

which yields

$$\sin \phi = \frac{2x}{1 + x^2}, \quad \cos \phi = \frac{1 - x^2}{1 + x^2}, \tag{5}$$

the expression (2) is rationalized in terms of x. If x is transcendental, we call ϕ *associated transcendental.*

Now we can state

THEOREM I. *The rotation group generated by the rotations* $A(\phi)$ *and* $B(\phi)$, *these two being rotations with rotation angle* ϕ *and with rotation axes perpendicular to each other, is a free (non-abelian) group (of rank two) for any fixed, associated transcendental value* ϕ.

Proof. It follows from Lemma II that any element (2) of the group H, generated by $A = A(\phi)$ and $B = B(\phi)$, is a non-constant matrix, if it cannot be transformed into identity by using the trivial relations. Now any non-trivial relation in H can be written in the form

$$A^{n_1} B^{m_1} A^{n_2} B^{m_2} \dots A^{n_k} B^{m_k} = I, \tag{6}$$

the product being a proper product if $k > 0$. Since (2) is a non-constant function of ϕ, (6) can be transformed, using (4) in a finite number of algebraic equations in x, not all vanishing identically. So substituting for ϕ any fixed associated transcendental number ϕ, no relation (6) is valid, and the theorem is proved.

Since in a free group generated by A and B the elements A^iBA^{-i} ($i = 0, 1, 2, \dots$) are free generators of a free group of infinite (but countable) rank, we get

COROLLARY. *For any fixed associated transcendental* ϕ *the rotations with rotation angle* ϕ *and with rotation axes in the same plane and making angles* $i\phi(i = 0, 1, 2, \dots)$ *with a fixed line in this plane are free generators of a free group (of infinite rank).*

4. Uncountable representations. J. von Neumann **(6)** proved that the set $\{x_t\} = M$ of distinct real numbers x_t, defined by

$$x_t = \sum_{n=0}^{\infty} 2^{2^{[nt]} - 2^{n^2}} \qquad (t > 0),$$

are algebraically independent over the field of rational numbers (no finite set $\{a_t\}$ of distinct numbers $a_t \in M$ satisfies an equation $P(y_t) = 0$, if $P(y_t)$ is a non-vanishing polynomial in the variables y_t with rational coefficients). Thus there are continuously many, distinct, associated transcendental numbers

$$\phi_t = 2 \arctan x_t \qquad (0 < t < 1). \tag{7}$$

Select another $\phi = \psi$ defined by $\psi = \phi_t$ with $t > 1$, t fixed.

Now we shall prove

THEOREM II. *The continuously many rotations* $R_t = A(\phi_t)\, B(\psi)\, A^{-1}(\phi_t)$ $(0 < t < 1)$ *are free generators of a free rotation group of continuous rank.*

We note that all R_t are rotations with the same rotation angle ψ and rotation axes in the same plane.[3] *In particular, the existence of a free rotation group of continuous rank has been established (without using the axiom of choice).*

Proof. The theorem is proved if any proper product $P(R_t)$ of a finite number of rotations R_t is unequal to the unity matrix. After simplifications we may write

$$P(R_t) = A(\phi_{t_1})\, B^{k_1}(\psi)\, A^{-1}(\phi_{t_1})\, A(\phi_{t_2})\, B^{k_2}(\psi)\, A^{-1}(\phi_{t_2}) \ldots A(\phi_{t_n})\, B^{k_n}(\psi)\, A^{-1}(\phi_{t_n}), \tag{8}$$

the k_i being integers $\neq 0$, the t_i $(i = 1, 2, \ldots, n)$ real numbers with

$$0 < t_i < 1, \quad t_i \neq t_{i+1} \qquad (i = 1, 2, \ldots, n-1).$$

Now replace in the right-hand side of (8) ψ by a real variable ϕ, and

$$\phi_{t_i}$$

by a multiplicity $m_i\phi$ of ϕ (m_i integers > 0) with $m_i = m_j$ if and only if $t_i = t_j$. After carrying out the simplifications

$$A^{-1}(m_i\,\phi)\, A(m_{i+1}\phi) = A(l_i\phi) = A^{l_i}(\phi) \qquad (l_i \neq 0),$$

(8) is transformed into a proper product (almost) of type (2), therefore—applying Lemma 2—into a non-constant matrix function of ϕ. From this it follows that at least one of the elements a_{ik} of matrix (8) is a non-constant function of

$$\phi_{t_i} \text{ and } \psi, \tag{9}$$

if we consider these for a moment as real variables. But then it is impossible —using substitution (7) and the result of von Neumann—that this function is equal to 0 or 1 if we substitute for the variables (9) their permitted associated transcendental and distinct values. So $P(R_t) \neq I$.

4.1. It is not necessary, of course, to take the rotation axes in the same plane to get free generators. The proof, just established, furnishes us a general method of generating free groups in the following way. Any rotation R can be written as a product

$$R = \begin{pmatrix} \cos\phi_u & -\sin\phi_u & 0 \\ \sin\phi_u & \cos\phi_u & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\phi_v & -\sin\phi_v \\ 0 & \sin\phi_v & \cos\phi_v \end{pmatrix} \begin{pmatrix} \cos\phi_w & -\sin\phi_w & 0 \\ \sin\phi_w & \cos\phi_w & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

[3]One might ask whether the geometrical structure of this set of rotation axes in the plane can be relatively simple, if we select a suitable set of continuously many values t. Indeed, it is possible that this set of rotation axes corresponding to the generating rotations is *perfect*. This follows easily from the fact that the set of numbers $\{x_t\}$ contains perfect subsets. To prove this, we observe that x_t is a monotonically increasing function of t; thus the set of transcendental numbers $\{x_t\}$ is nowhere dense in the set of all real numbers, and is, moreover, a G_δ-set; therefore it contains perfect subsets.

by electing suitable ϕ_u, ϕ_v, ϕ_w (Eulerian angles, see (**1**, p. 104)). Let the u, v, w range as real variables over certain sets, say $0 < u < 1, 1 < v < 2, 2 < w < 3$.

Now we consider elements R corresponding with triplets (u, v, w) differing from each other in each of the variables u, v, w. Then the elements are free generators of a free group. Indeed, a version of Lemma II on any proper product can be applied after simplifications.

Briefly the proper products do not vanish identically as functions, and cannot therefore be equal to unity for permitted values of their variables, since these values are algebraically independent.

4.2. Remarks. If we consider $A(\phi)$ and $B(\phi)$ as matrix functions (the elements being analytic functions of the real variable ϕ) it follows from Lemma II and Theorem I that these matrix functions are free generators of a free group (the only constant function in the group being the unity matrix).

In a certain analogy with the generators of Theorem II, one can also consider the family of matrix functions

$$A(\phi_\alpha)\, B(\psi)\, A^{-1}(\phi_\alpha), \tag{10}$$

the indices α ranging over a set of arbitrary potency m (ψ and ϕ_α being real variables). Therefore *these orthogonal matrix functions* (9) *are free generators of a free group of rank m. Any* free group can therefore be represented isomorphically by a system of orthogonal matrix functions. Perhaps this may be of some use for the theory of free groups.

5. Conjectures. One could try to prove Theorem I by the alternative method of expanding $\sin \phi$ and $\cos \phi$ in a Taylor series.

Writing

$$\sin x = x + o(x^2), \quad \cos x = 1 - \tfrac{1}{2}x^2 + o(x^2),$$

one sees easily that for small x all products of type (2) with $k \leqslant 2$ give a non-constant matrix. But this fails already in the case $k = 3$; taking, for example

$$n_1 = 2, n_2 = 3, n_3 = -5, \quad m_1 = -5, m_2 = 2, m_3 = 3.$$

However, for $k \leqslant 3$, a proof is possible if we expand $\sin x$ and $\cos x$ up to $o(x^3)$. It will perhaps be possible to get a proof of Theorem I using induction; however, the computations involved are very lengthy. On the other hand, it may be possible to generalize this method in cases where the generating rotations A and B are not perpendicular to each other. Consider

$$A' = CAC^{-1} \text{ with } C = B(\alpha)$$

for a fixed but arbitrary α and carry out the computations for products (2), in which B is replaced by A'. This gives the following *conjecture* (generalizing Theorem I): two rotations with arbitrary but different rotation axes are free

generators of a free group for all rotation angles ϕ, a countable number of values ϕ excepted.

We conclude with another *conjecture*: Let $\{\mathfrak{G}_\alpha\}$ be a family of at most continuously many groups $\mathfrak{G}_\alpha$, each of which is countable (or more generally consists of less than continuously many elements) and can be represented as a three-dimensional rotation group; now the free product of the groups $\mathfrak{G}_\alpha$ can be isomorphically represented by a rotation group.

References

1. L. Bieberbach, *Analytische Geometrie* (Teubner, 1932).
2. Kh. A. Doniakhi, *Linear representation of the free product of cyclic groups*, Leningrad State Univ. Annals [Uchenye Zapiski] Math. Ser., *10* (1940), 158–165 (Russian).
3. D. J. Fuchs-Rabinowitsch, Leningrad State Univ. Annals [Uchenye Zapiski] Math. Ser., *10* (1940), 154–157 (Russian).
4. J. de Groot and T. Dekker, *Free subgroups of the orthogonal group*, Comp. Math., *12* (1954), 134–136.
5. F. Hausdorff, *Grundzüge der Mengenlehre* (1914), 469–472.
6. J. von Neumann, *Ein System algebraisch unabhängiger Zahlen*, Math. Ann., *99* (1928), 134–141.
7. R. M. Robinson, *On the decomposition of spheres*, Fund. Math., *34* (1947), 246–260.
8. I. N. Sanov, *A property of a representation of a free group*, Doklady Akad. Nauk. SSSR (N.S.), *57* (1947), 657–659 (Russian).
9. W. Sierpiński, *Sur le paradoxe de la sphère*, Fund. Math., *33* (1945), 235–244.

University of Amsterdam

ON A THEOREM OF BAER AND HIGMAN

SEÁN TOBIN

1. Introduction

1.1 Baer has shown **(1)** that if the fact that the exponent of a group is m (that is, m is the least common multiple of the periods of the elements) implies a limitation on the class of the group, then m must be a prime. Graham Higman has extended this result by proving **(3)** that for any given integer M there are at most a finite number of prime powers q other than primes, such that the fact that a group has exponent q implies a limitation on the class of the Mth derived subgroup. In fact, given arbitrary positive integers M and N, he produces, by an intricate construction, a finite group G having derived length $M + 2$ and prime-power exponent p^r, such that the class of the Mth derived subgroup of G exceeds N, where

$$p^r - 1 > (p - 1)\,A(M)$$

and $A(M)$ is an integer-valued function:

$$A(0) = 1, \quad A(1) = 3, \quad A(2) = 13, \ldots.$$

The case $M = 0$ is Baer's result.

1.2 In this paper we consider those prime powers p^r for which

$$p^r - 1 = (p - 1)\,A(M),$$

in the special cases $M = 0$ and $M = 1$, that is, $r = 1$ and $p = r = 2$ respectively. We show that no result similar to that of Higman can be obtained in these cases; indeed, an upper bound is given for the class of the Mth derived subgroup in terms of the derived length.

Specifically, the final results are as follows:

THEOREM 1. *If a finitely generated group G has exponent 4 and ϕ-length λ, then the class of $\phi(G)$ is at least $2^{\lambda-2}$ and at most $5^{\lambda-2}$.*

The meaning of $\phi(G)$ and ϕ-length of G is explained in §2.1.

THEOREM 2. *If a finitely generated group G has prime exponent p and derived length d, then the class of G is at least 2^{d-1} and at most p^{d-1}.*

Received July 22, 1955. This paper embodies some of the work carried out by the author for a Ph.D. thesis at the University of Manchester. I wish to express my gratitude to Dr. Graham Higman for his continued interest and advice then, and, in connection with the present paper, for Theorem 5.2 which is essentially due to him both in conception and in proof.

A well-known result due to Hall (2) gives $2^{\lambda-2}$ and 2^{d-1}, respectively, as the lower bounds; we shall be concerned here with the upper bounds. The interest of these results lies, of course, not so much in the bounds given for the class, which are presumably far from best possible if λ and d are large, as in the fact that bounds exist which are independent of the number of generators of the groups in question.

We may mention also an auxiliary theorem which is of interest in itself.

THEOREM 5.2. *If, for a finitely generated group G with prime-power exponent p^r, there exists a positive integer s such that*

$$H_{s+1}(G) \subseteq D_2(G),$$

then

$$H_{qs+1}(G) \subseteq H_{q+1}(D(G)), \qquad q = 1, 2, 3, \ldots .$$

Here $H_i(G)$ and $D_2(G)$ are members of the lower central series and the derived series (§2) of G, respectively.

2. Definitions

2.1 The *Frattini subgroup* $\phi(G)$ of a group G is defined to be the intersection of all the maximal subgroups of G; if P is a p-group (by which we mean that the order of P is a power of a prime p) it is known (2) that $\phi(P) = D(P) \cup P^p$ where $D(P)$ is the commutator subgroup and P^p the subgroup generated by the pth powers of all elements in P.

The *Frattini series* is defined inductively:

$$\phi_0(G) = G, \qquad \phi_{i+1}(G) = \phi(\phi_i(G)), \qquad i > 0.$$

If this series terminates with the identity (as it certainly does for a finite group), so that $\phi_{j-1}(G) \neq \{1\}$ but $\phi_i(G) = \{1\}$ whenever $i > j$, we shall say that the *ϕ-length of G* is j.

The *derived series* of a group G is defined inductively:

$$D_0(G) = G; \quad D_{i+1}(G) = D(D_i(G)), \qquad i > 0.$$

If N is a normal subgroup of a p-group P, such that $\phi_i(P) \supseteq N$, it is easily seen that $\phi_i(P/N) = \phi_i(P)/N$.

Since $u^{-1}v^{-1}uv = (u^{-1})^2(uv^{-1})^2v^2$, we see that if P is a 2-group, then $P^2 \supseteq D(P)$ and $\phi(P) = P^2$. If, in addition, P is generated by elements which all have period 2, then $D(P) \supseteq P^2$, and consequently $\phi(P) = D(P) = P^2$.

Again, if P is a 2-group with n independent generators, $P/\phi(P)$ is elementary abelian with order 2^n; thus every factor-group $\phi_i(P)/\phi_{i+1}(P)$ in the ϕ-chain of P is an elementary abelian 2-group (i.e., the direct product of cycles of order 2).

2.2 Square brackets will be used to denote commutation:

$$[x, y] = x^{-1}y^{-1}xy.$$

If $x_1, x_2, x_3, \ldots$ are arbitrary elements in a group, the (complex) commutators in the x_i are defined inductively by the rules

(i) x_i is a commutator;

(ii) if c and d are commutators in these elements, so also is $[c, d]$. In particular, a *left-normed* (or simple) *n-fold commutator* is defined:

$$[x_1, x_2, \ldots, x_n] = [[x_1, x_2, \ldots, x_{n-1}], x_n] \qquad (n > 2).$$

The *weight* of a commutator *in the element* x_i is defined by

(i) the weight of x_j in x_i is 1 if $i = j$, 0 if $i \neq j$.

(ii) the weight of $[c, d]$ in x_i is the sum of the weights of c and d in x_i.

The *weight of a commutator* is the sum of its weights in the components x_i. We recall the commutator identity

$$[x, yz] = [x, z][x, y][x, y, z].$$

The *lower central series* of a group G is defined inductively

$$H_1(G) = G; \quad H_{i+1}(G) = [H_i(G), G], \qquad i > 1.$$

(That is, H_{i+1} is the subgroup generated by the set of all commutators of the form $[h_i, g]$ with h_i in H_i and g in G.) A discussion of the properties of this series, and its connection with the class of G, may be found in **(2)**.

3. A certain group with exponent 4

3.1 Let G be a 2-group with ϕ-length 3. Let $H = \phi_2(G)$, and let $K \simeq G/H$. It is clear that the exponent of G must be 8 or 4, and the exponent of K is 4. Both H and $\phi(K)$ are non-trivial elementary abelian 2-groups, and $\phi(K) \simeq \phi(G)/H$. We use 1 to represent the unit element of G or K according to context.

In this section we show that the requirement that G have exponent 4 introduces a certain relation into the group ring of K over the field of two elements; and in §3.2 we show that this relation yields an upper bound for the class of $\phi(G)$.

For each element k in K we choose, in the corresponding coset of G/H, a coset representative g_k in G. If x, y are any elements of K, the element $h(x, y)$ in H is determined by the equation

$$g_x g_y = g_z h(x, y) \qquad \text{where } z = xy.$$

The element g_k induces an automorphism of H which depends only on k; if h is any element of H we denote its image by

$$h^k = g_k^{-1}\, h\, g_k.$$

The automorphisms k belong to the ring Θ of endomorphisms of H, and Θ has characteristic 2. Since $(h^x)^y = h^z$, where x, y and z have the meanings already assigned, the subring of Θ generated by the set $\{k : k \in K\}$ is a homomorphic image of the group ring of K over the field of two elements (0, 1).

Any element g of G can be written uniquely in the form $g = g_k h$ with k in K and h in H. Then

$$g^4 = g_1 h(k^2, k^2)\, h(k, k)^{(1+k)^2}\, h^{(1+k)^3}.$$

We choose $g_1 = 1$; and now in order that G itself may have exponent 4 it is necessary and sufficient that

(i) $h^{(1+k)^3} = 1$ for all choices of h in H and k in K

and

(ii) $h(k^2, k^2)\, h(k, k)^{(1+k)^2} = 1$ for all choices of k in K.

We shall consider condition (i), which is more amenable to treatment than (ii). Since

$$h^{(1+k)} = hg_k^{-1}hg_k = [h, g_k],$$

what relation (i) says, in effect, is that

$$[h, u, v, w] = 1$$

for any element h in H, whenever the elements u, v, w all lie in the same coset of G modulo H.

3.2 Thus we consider the group ring of K over the field of two elements $(0, 1)$, with the relation $(1 + k)^3 = 0$ for all elements k in K.

We use the notation $K_i = 1 + k_i$ for elements k_i in K; throughout what follows we shall *not* use k without a subscript to represent an element of K. To avoid repetition, we make the following convention:

k_1 is an arbitrary element of K

$k_2 = k_1^2$

k_3, k_5, k_7 are arbitrary elements of $\phi(K)$

$k_4 = [k_1, k_3]$, $k_6 = [k_1, k_5]$, $k_8 = [k_1, k_7]$.

Thus $K_i^2 = 0$ and $K_i K_j = K_j K_i$ when i and j lie between 2 and 8 inclusive.

The relation in the group ring may now be written

3.21 $$K_1 K_2 = 0.$$

If we replace k_1 here by $k_1 k_3$, then k_2 is replaced by $(k_1 k_3)^2 = k_2 k_4$,

$$K_1 \quad \text{becomes} \quad 1 + k_1 k_3 = 1 + (1 + K_1)(1 + K_3) = K_1 + K_3 + K_1 K_3,$$

and the relation gives

3.22 $$(K_1 + K_3 + K_1 K_3)(K_2 + K_4 + K_2 K_4) = 0.$$

Post-multiplication by K_4 gives $(K_1 + K_3 + K_1 K_3) K_2 K_4 = 0$; then post-multiplication by K_2 gives

3.23 $$K_1 K_3 (K_2 + K_4) = 0.$$

This leaves $(K_1 + K_3)(K_2 + K_4) = 0$, but $K_1 K_2 = 0$, thus finally

3.24 $$K_1 K_4 + K_3 (K_2 + K_4) = 0.$$

Using 3.23, premultiplication by K_1 gives

3.25 $$K_2 K_4 = 0.$$

In 3.24, replace k_3 by k_3k_5, then k_4 is replaced by $[k_1, k_3k_5] = k_4k_6$. Thus

3.26 $$K_3K_6 + K_4K_5 + (K_1 + K_3 + K_5 + K_3K_5)K_4K_6 + K_2K_5(K_2 + K_4 + K_6 + K_4K_6) = 0.$$

If we substitute for k_5 the particular value $k_5 = [k_1, k_\beta]$ where k_β is an arbitrary element in $\phi(K)$, then

$$k_6 = [k_\beta, k_1, k_1] = [k_\beta, k_1^2] = 1$$

and $K_2K_5 = 0$ by 3.25, thus

$$K_4K_5 + K_3K_4K_5 = 0, \quad \text{which implies} \quad K_4K_5 = 0.$$

Consequently, in equation 3.26, where k_5 is again an arbitrary element of $\phi(K)$, we have

3.27 $$K_4K_6 = 0$$

and 3.26 reduces to

3.28 $$K_3K_6 + K_4K_5 + K_2K_3K_5 = 0.$$

Again, in 3.28 replace k_5 by k_5k_7 and k_6 by k_6k_8. Then, since $K_6K_8 = 0$ by 3.27

$$K_3(K_6 + K_8) + K_4(K_5 + K_7 + K_5K_7) + K_2K_3(K_5 + K_7 + K_5K_7) = 0.$$

Using 3.28 this simplifies to

$$K_4K_5K_7 = K_2K_3K_5K_7.$$

On multiplying 3.28 by K_7 and using this, we obtain $K_3K_6K_7 = 0$, which is equivalent to saying $K_4K_5K_7 = 0$. Consequently

3.29 $$K_2K_3K_5K_7 = 0.$$

Now if we take any element k_9 in $\phi(K)$, then

$$K_9 = K_{11}^2 + K_{12}^2 + \ldots + K_{1\xi}^2 + \text{(products of two or more of these squares)}$$

where $k_{11}, k_{12}, \ldots, k_{1\xi}$ are certain elements of K. Thus 3.29 implies that

$$K_3K_5K_7K_9 = 0$$

for all choices of four elements k_3, k_5, k_7, k_9 in $\phi(K)$.

In the group G this means that for arbitrary elements w, x, y, z in $\phi(G)$ and h in H, $[h, w, x, y, z] = 1$. Thus we have proved

THEOREM 3.2. *If G is a finite group with exponent 4 and ϕ-length 3, then the class of $\phi(G)$ is at most 5.*

3.3 Let G be a finitely generated group with exponent 4: such a group is finite **(5)**, hence a 2-group. Thus $\phi(G/\phi_3(G)) = \phi(G)/\phi_3(G)$ and the derived series of $\phi(G)$ coincides with its Frattini series. Thus from Theorem 3.2 we obtain, using the notation previously explained,

COROLLARY 3.3. *If G is a finitely generated group with exponent 4,*

$$H_6(\phi(G)) \subseteq D_2(\phi(G)).$$

4. A similar result

Meier-Wunderli has shown, in **(4)**, that a finitely generated metabelian group with prime exponent p has class at most p. This may be stated as follows:

THEOREM 4.1 (Meier-Wunderli). *If G is a finitely generated group with prime exponent p,*

$$H_{p+1}(G) \subseteq D_2(G).$$

This result bears an obvious resemblance to Corollary 3.3; they will be extended simultaneously by means of the theorems given in the next section.

5. Some theorems on commutators

5.1 LEMMA. *Let $x_1, x_2, \ldots, x_n$ be arbitrary elements of an arbitrary group G; let c be a commutator of positive weight in each of the x_i ($i = 1, 2, \ldots, n$) and let the equation*

$$c = d_1 d_2 \ldots d_a,$$

where the d_j are also commutators in $x_1, x_2, \ldots, x_n$, be an identity in the group variables x_i. Then there is an equation

$$c = b_1 b_2 \ldots b_\beta$$

also true for all $x_1, x_2, \ldots, x_n$ in G, where the b's are commutators in the elements $d_1, d_2, \ldots, d_a$ such that every b is of positive weight in each of $x_1, x_2, \ldots, x_n$.

Proof. This can be proved by induction on n, being trivially true for $n = 1$. Thus we may suppose that each d_j is of positive weight in each of $x_1, \ldots, x_{n-1}$. We may further suppose that the commutators of zero weight in x_n are those in an initial segment $d_1 d_2 \ldots d_\xi$. For if this is not so, then using the relation $yx = xy[y, x]$ we can bring them to the left of the expression one at a time, by a process which terminates, since the new commutators $[y, x]$ introduced have positive weight in x_n.

When this has been done, let $x_n = 1$. Then every commutator of positive weight in x_n reduces to the identity, while those of zero weight in x_n are not affected. Thus for all $x_1, x_2, \ldots, x_n$ in G

$$1 = d_1 d_2 \ldots d_\xi.$$

Hence also, for all $x_1, x_2, \ldots, x_n$ in G

$$c = d_{\xi+1} d_{\xi+2} \ldots d_a$$

which is the expression required.

5.2 THEOREM. *If, for a finitely generated group G with prime-power exponent p^r, there exists a positive integer s such that*

$$H_{s+1}(G) \subseteq D_2(G),$$

then

$$H_{qs+1}(G) \subseteq H_{q+1}(D(G)), \qquad q = 1, 2, 3, \ldots.$$

Before proving Theorem 5.2, we make a remark which also has a bearing on Theorem 6.2. A well-known theorem, due to O. Schreier, states that in a finitely generated free group any subgroup of finite index is also finitely generated. Since any finitely generated group is a homomorphic image of a finitely generated free group, the statement remains true when the words "free group" are replaced by "group." In particular, let F be a group with a finite exponent. Then if F is finitely generated, so also is $D(F)$ and every factor $D_i(F)/D_{i+1}(F)$ in the derived series is finite.

Proof. If N is any normal subgroup of G, $H_i(G/N) = \{H_i(G), N\}/N$; hence it is sufficient to prove

$$H_{qs+1}(X) \subseteq H_{q+1}(D(X))$$

for the finite group $X = G/D_\alpha(G)$ where α is chosen large enough to ensure that $D_\alpha(G) \subseteq H_{q+1}(D(G))$.

Thus we consider a p-group X of exponent p^r, generated by a minimal basis $x_1, x_2, \ldots, x_n$. A result due to Hall (**2**; Theorem 2.8.2) states that $H_j(X)$ is generated by the set of all left-normed commutators of weight $\geq j$ in the components $x_1, x_2, \ldots, x_n$. Thus $D(X) = H_2(X)$ is generated by the simple commutators

$$[x_{i_1}, x_{i_2}, \ldots, x_{i_t}] \qquad t \geq 2.$$

But $H_{s+1}(X) \subseteq D_2(X)$; hence the simple commutators with $t \geq s + 1$ all lie in $D_2(X) \subseteq \phi(D(X))$, and can therefore (**2**) be omitted from any generating set of $D(X)$. Thus $D(X)$ is in fact generated by the simple commutators with $2 \leq t \leq s$; we denote these by $d_1, d_2, \ldots$.

If we set

$$c = [x_{i_1}, x_{i_2}, \ldots, x_{i_{qs+1}}],$$

then c lies in $D_2(X)$, so

$$c = d_{j_1} d_{j_2} \ldots d_{j_m}.$$

Consequently, by Lemma 5.1,

$$c = b_1 b_2 \ldots b_\beta,$$

where each b_i is a commutator in the d's and therefore also in the x's and is of positive weight in x_{i_ν}, $\nu = 1, 2, \ldots, qs + 1$.

Thus, in particular, each b_i is of weight at least $qs + 1$ in the x's. Since no commutator d has weight greater than s in the x's, each b_i must be of weight at least $q + 1$ in the d's. The required result follows.

5.3 Corollary. *If, in addition to the assumptions of the previous theorem,*

$$H_{s+1}(D_\beta(G)) \subseteq D_{\beta+2}(G) \text{ for every positive integer } \beta,$$

then

$$H_{s^d+1}(G) \subseteq D_{d+1}(G) \text{ for any positive integer } d.$$

Proof. $H_{s+1}(G) \subseteq D_2(G)$ gives a basis for induction on d. We assume that the statement is true for an integer d. Then

$$\begin{aligned} H_{s^{1+d}+1}(G) &\subseteq H_{s^d+1}(D(G)) && \text{by Theorem 5.2,} \\ &\subseteq D_{d+1}(D(G)) && \text{by the induction hypothesis,} \end{aligned}$$

since $D(G)$ itself satisfies the conditions stated for G.

6. Final results

All that remains now is to apply 5.3 to the groups considered in 3.3 and 4.1. If G is finitely generated with exponent 4, so also are its successive Frattini subgroups; and

$$D_\beta(\phi(G)) = \phi_\beta(\phi(G)) = \phi(\phi_\beta(G));$$

thus by Corollary 3.3,

$$H_{8+1}(D_\beta(\phi(G)) \subseteq D_{\beta+2}(\phi(G)).$$

Corollary 5.3 now gives

THEOREM 6.1. *If G is a finitely generated group with exponent* 4:

$$H_{8^\lambda+1}(\phi(G)) \subseteq \phi_{\lambda+2}(G).$$

Again, if G is finitely generated with prime exponent p, so also is $D_\beta(G)$, and by Theorem 4.1,

$$H_{p+1}(D_\beta(G)) \subseteq D_{\beta+2}(G).$$

THEOREM 6.2. *If G is a finitely generated group with prime exponent p*:

$$H_{p^d+1}(G) \subseteq D_{d+1}(G).$$

REFERENCES

1. R. Baer, *The higher commutator subgroups of a group.*, Bull. Amer. Math. Soc., *50* (1944), 143–160.
2. P. Hall, *A contribution to the theory of groups of prime power order*, Proc. Lond. Math. Soc., *36* (1933), 29–95.
3. G. Higman, *Note on a theorem of R. Baer*, Proc. Camb. Phil. Soc., *45* (1949), 321–327.
4. H. Meier-Wunderli, *Über endliche p-Gruppen deren Elemente der Gleichung $x^p = 1$ genügen*, Comment. Math. Helv., *24* (1950), 18–45.
5. I. N. Sanov, *Solution of Burnside's Problem for exponent 4*, Leningrad State Univ. Annals, Math. Ser., *10* (1940), 166–170.

University College, Galway, Ireland

ON COMMUTING RINGS OF ENDOMORPHISMS

C. W. CURTIS

1. Introduction. Various problems concerning the general theory of centralizers of modules which are not assumed to be completely reducible have been discussed by Fitting **(3)**, Brauer **(2)**, and Nakayama. In this paper we present a new approach to some of these questions, which has its origin in Weyl's discussion **(15)** of the centralizer of a finite group of collineations.

Let $\mathfrak{B}$ be a ring with an identity element, and let $\mathfrak{M}'$ and $\mathfrak{M}$ be unital[1] left and right $\mathfrak{B}$-modules, respectively. We assume that there exists a function $\tau(\psi, x)$ on $\mathfrak{M}' \times \mathfrak{M} \rightarrow \mathfrak{B}$ which is bilinear with respect to $\mathfrak{B}$, and non-degenerate. The set $\mathfrak{b}$ of all finite sums $\sum \tau(\psi_i, x_i)$ is a two-sided ideal in $\mathfrak{B}$, called the nucleus of the pairing $(\mathfrak{M}', \mathfrak{M}, \tau)$. Let $\mathfrak{E}$ be the ring of all $\mathfrak{B}$-endomorphisms of $\mathfrak{M}$. Then $\mathfrak{E}$ contains the right ideal $\mathfrak{M}' \odot \mathfrak{M}$ consisting of all finite sums of the endomorphisms $\psi \odot u$ of $\mathfrak{M}$, where $x(\psi \odot u) = u\tau(\psi, x)$, $x \in \mathfrak{M}$. By a centralizer $\mathfrak{C}$ of $\mathfrak{M}$ relative to $\mathfrak{B}$ we mean a subring $\mathfrak{C}$ of $\mathfrak{E}$ containing the right ideal $\mathfrak{M}' \odot \mathfrak{M}$.

Our basic assumption is that the nucleus $\mathfrak{b}$ contain a two-sided identity element. Then it is proved in §5 that the ring of $\mathfrak{C}$-endomorphisms of $\mathfrak{M}$ is precisely the set of endomorphisms $R_b: x \rightarrow xb$ determined by the elements of $\mathfrak{B}$. Let $\mathfrak{N}$ be a $\mathfrak{C}$-direct summand of $\mathfrak{M}$; then $\tau(\mathfrak{M}', \mathfrak{N})$ is a left ideal in $\mathfrak{b}$, and the mapping $\mathfrak{N} \rightarrow \tau(\mathfrak{M}', \mathfrak{N})$ is a (1–1) mapping, preserving direct sums, intersections, and isomorphism relations, between the set of $\mathfrak{C}$-direct summands of $\mathfrak{M}$ and the set of left ideal direct components of $\mathfrak{b}$. Dually, if $\mathfrak{M}' \odot \mathfrak{M}$ contains the identity operator on $\mathfrak{M}$, and if the pairing $\psi \odot u$ is non-degenerate, then the mapping $\mathfrak{N} \rightarrow \mathfrak{M}' \odot \mathfrak{N}$ defines a (1–1) mapping between the set of $\mathfrak{B}$-direct summands of $\mathfrak{M}$ and the set of left ideal direct components of the centralizer $\mathfrak{C}$. If $\mathfrak{B}$ satisfies the minimum condition for left ideals, then every indecomposable $\mathfrak{C}$-direct summand $\mathfrak{N}$ of $\mathfrak{M}$ contains a unique maximal $\mathfrak{C}$-submodule, and if $\mathfrak{N}_1$ and $\mathfrak{N}_2$ are indecomposable $\mathfrak{C}$-direct summands, then $\mathfrak{N}_1$ and $\mathfrak{N}_2$ are $\mathfrak{C}$-isomorphic if and only if $\mathfrak{N}_1/\mathfrak{S}_1$ and $\mathfrak{N}_2/\mathfrak{S}_2$ are $\mathfrak{C}$-isomorphic, where $\mathfrak{S}_i$ is the unique maximal submodule of $\mathfrak{N}_i$, $i = 1, 2$.

The principal application of this theory is to projective (or ray) representations of a finite group $\mathfrak{G}$ by s.l.t. (semi-linear transformations) of a vector space $\mathfrak{M}$ over a division ring Δ. If $\mathfrak{B} = \Delta(\mathfrak{G}, H, \rho)$ is the crossed product associated with the projective representation, then it is proved in §2 that a space $\mathfrak{M}'$, and a pairing τ of $\mathfrak{M}' \times \mathfrak{M} \rightarrow \mathfrak{B}$ which satisfies our hypotheses,

Received August 27, 1955. Presented to the American Mathematical Society August 30, 1955. (The author is a National Research Fellow.)

[1]A left or right $\mathfrak{B}$-module $\mathfrak{M}$ is called *unital* if the identity element of $\mathfrak{B}$ acts as identity operator on $\mathfrak{M}$.

can be constructed if and only if the normalized factor set ρ satisfies the condition $\rho_{s,s}^{-1} = 1$ for all s in $\mathfrak{G}$. In §3 the pairing considered by Weyl **(15)** is defined, and shown to satisfy our hypotheses, so that Weyl's results are consequences of the theorems proved in §5. In §4 and §8 some special results are derived which concern the pairings obtained in §2 from projective representations of finite groups. A few remarks are included in §9 on the applications of the results on projective representations to the Galois theory of primitive rings with minimal ideals. A direct proof is given in §10 of the fact that the centralizer of a symmetric algebra $\mathfrak{A}$ of l.t. in a finite dimensional vector space $\mathfrak{M}$ which is a projective $\mathfrak{A}$-module is a symmetric algebra.

2. Projective representations of finite groups[2]. Let $\mathfrak{M}$ be a commutative group, and Δ a division ring consisting of endomorphisms $\xi: x \to x\xi$ of $\mathfrak{M}$, such that Δ contains the identity mapping. Then $\mathfrak{M}$ is a right vector space over Δ. Two non-singular s.l.t. T_1 and T_2 in $\mathfrak{M}$ over Δ are said to be equivalent if $T_1 = T_2\mu$, where μ is a non-zero element of Δ. An equivalence class $\{T\}$ of non-singular s.l.t. is called a *projective transformation*. Multiplication of projective transformations is defined in the obvious way, and the projective transformations form a group $\mathfrak{P}(\mathfrak{M}, \Delta)$.

Now let $\mathfrak{G} = \{1, s, t, \ldots\}$ be a finite group. A homomorphism of $\mathfrak{G}$ into $\mathfrak{P}(\mathfrak{M}, \Delta)$ is called a *projective representation* of $\mathfrak{G}$. Evidently a projective representation is determined by a mapping $s \to T_s$ of $\mathfrak{G}$ into the set of non-singular s.l.t. of $\mathfrak{M}$ such that

$$T_sT_t = T_{st}\,\rho_{s,t}, \tag{1}$$

where the $\rho_{s,t}$ are certain non-zero elements of Δ. From the associative law and (1) we obtain

$$\xi^{\bar{s}\,\bar{t}} = \rho_{s,t}^{-1}\,\xi^{\overline{st}}\,\rho_{s,t}, \tag{2}$$

where $\bar{s}: \xi \to \xi^{\bar{s}}$ and $\bar{t}$ are the automorphisms of Δ determined by the s.l.t. T_s and T_t, and

$$\rho_{s,tu}\,\rho_{t,u} = \rho_{st,u}\,\rho_{s,t}^{\bar{u}}. \tag{3}$$

If we denote the inner automorphism $\xi \to \rho_{s,t}^{-1}\xi\rho_{s,t}$ by $\bar{\rho}_{s,t}$, then (2) becomes

$$\bar{s}\,\bar{t} = \overline{st}\,\bar{\rho}_{s,t}. \tag{2'}$$

A set $\{\rho_{s,t}; \bar{u}\}$, where the $\rho_{s,t}$ are non-zero elements of Δ, and the $\bar{u}$ are automorphisms of Δ, is called a *factor set* of $\mathfrak{G}$ (in Δ) if the equations (2) and (3) hold. Thus the transformation T_s satisfying (1) determine a factor set $\{\rho_{s,t}; \bar{u}\}$. If we replace the representatives T_s of the projective transformations corresponding to the elements of $\mathfrak{G}$ by new representatives $T'_s = T_s\mu_s$, then we obtain

$$T'_sT'_t = T'_{st}\rho'_{s,t}$$

[2]For the terminology introduced in the first part of this section, see **(6**, Chap. 4, §17, 18**)**.

where the automorphisms $\bar{s}'$, associated with T', satisfy

(4) $$\bar{s}' = \bar{s}\,\bar{m}_s,$$

where $\bar{m}_s$ is the inner automorphism $\xi \rightarrow \mu_s^{-1}\xi\mu_s$, and

(5) $$\rho'_{s,t} = \mu_{st}^{-1}\,\rho_{s,t}\,\mu_s^{\bar{t}}\,\mu_t.$$

Thus it is natural to say that two factor sets $\{\rho_{s,t}; \bar{u}\}$ and $\{\rho'_{s,t}; \bar{u}'\}$ are equivalent if there exist elements $\mu_s \neq 0$ such that (4) and (5) hold. Then a projective representation determines a class of equivalent factor sets.

Now let $\{\rho_{s,t}; \bar{u}\}$ be a factor set, and let $\{b_s\}$ be a set of elements in (1–1) correspondence with the elements in $\mathfrak{G}$. The set $\mathfrak{B}$ of formal expressions $\sum b_s\xi_s$, $\xi_s \in \Delta$, $s \in \mathfrak{G}$ becomes an associative ring if we define two expressions to be equal if and only if they have the same coefficients, and if addition is defined componentwise, and multiplication using the distributive laws and the rules

$$b_s b_t = b_{st}\,\rho_{s,t},$$
$$\xi b_s = b_s\,\xi^{\bar{s}}.$$

Then $\mathfrak{B}$ is called a *crossed product* $\Delta(\mathfrak{G}, H, \rho)$ with correspondence $s \rightarrow \bar{s} = s^H$, and factor set ρ. If $\{\rho'_{s,t}; \bar{u}'\}$ is a factor set equivalent to $\{\rho_{s,t}; \bar{u}\}$, and if $\mathfrak{B}'$ is a crossed product $\Delta(\mathfrak{G}, H', \rho')$ with correspondence

$$s \rightarrow s^{H'} = \bar{s}$$

and factor set ρ', then it is easily verified that $\mathfrak{B}$ and $\mathfrak{B}'$ are isomorphic.

As Jacobson observes (**6**, p. 82), the element $b_1\rho_{1,1}^{-1}$ is an identity 1 for $\mathfrak{B}$, and if we identify Δ with the division subring 1Δ of $\mathfrak{B}$, then every element of $\mathfrak{B}$ can be expressed uniquely in the form $\sum b_s\xi_s$, where the term $b_s\xi_s$ is now the product of $b_s = b_s 1$ with ξ_s. It follows that $\mathfrak{B}$ is a two-sided vector space of finite left and right dimension over Δ, and consequently $\mathfrak{B}$ satisfies both chain conditions for left and right ideals.

If $s \rightarrow T_s$ defines a projective representation of $\mathfrak{G}$ with correspondence H and factor set ρ, then

$$\sum b_s\xi_s \rightarrow \sum T_s\xi_s$$

defines a representation of $\mathfrak{B}$ by endomorphisms of the representation space $\mathfrak{M}$ such that the identity element of $\mathfrak{B}$ is mapped onto the identity mapping in $\mathfrak{M}$, while conversely any such representation of $\mathfrak{B}$ by endomorphisms of $\mathfrak{M}$ gives rise to a projective representation of $\mathfrak{G}$ with the same correspondence and factor set.

Now let $\mathfrak{M}$ be a unital right $\mathfrak{B}$-module, and hence, in particular, a right vector space over Δ. Let $\mathfrak{M}'$ be a left vector space dual to $\mathfrak{M}$ with respect to a non-degenerate bilinear form $\langle\psi, x\rangle$ on $\mathfrak{M}' \times \mathfrak{M} \rightarrow \Delta$, such that the s.l.t. R_s: $x \rightarrow xb_s$ determined by the elements of $\mathfrak{G}$ all have transposes R_s^* relative to the form $\langle\psi, x\rangle$. Thus R_s^* is a s.l.t. of $\mathfrak{M}'$ with automorphism $\bar{s}^{-1}$ such that

(if we write operators on $\mathfrak{M}'$ to the left),

$$\langle \psi, xR_s\rangle^{\bar{s}^{-1}} = \langle R_s^{*}\, \psi, x\rangle \tag{6}$$

for all ψ and x.

We prove first that if we set $(\sum b_s\xi_s)\,\psi = \sum R_s^*(\xi_s\psi)$, then $\mathfrak{M}'$ becomes a unital left $\mathfrak{B}$-module. For all x and ψ, we have, since $1 = b_1\rho_{1,1}^{-1}$,

$$\langle 1\psi, x\rangle = \langle R_1^{*}(\rho_{1,1}^{-1}\psi), x\rangle = \langle \rho_{1,1}^{-1}\psi, xR_1\rangle^{\bar{1}^{-1}} = \langle \psi, x\rangle$$

by (2′), and hence $1\psi = \psi$. In order to prove that $\mathfrak{M}'$ is a left $\mathfrak{B}$-module, it is sufficient to prove that $(ab)\psi = a(b\psi)$. For all x and ψ, we have

$$\begin{aligned}\langle (b_s\xi b_t\eta)\,\psi, x\rangle &= \langle R_{st}^{*}(\rho_{s,t}\xi^{\bar{t}}\eta\psi), x\rangle\\ &= \langle \rho_{s,t}\xi^{\bar{t}}\eta\psi, xR_{st}\rangle^{\overline{st}^{-1}} = \langle \xi^{\bar{t}}\eta\psi, xR_sR_t\rangle^{\bar{\rho}_{s,t}^{-1}\overline{st}^{-1}}\\ &= \langle R_s^{*}(R_t^{*}(\xi^{\bar{t}}\eta\psi)), x\rangle^{\bar{s}\,\bar{t}\bar{\rho}_{s,t}^{-1}\overline{st}^{-1}} = \langle b_s\xi(b_t\eta\psi), x\rangle\end{aligned}$$

by (2′), and the conclusion follows from the non-degeneracy of the form.

We wish to study the centralizer of $\mathfrak{M}$ relative to $\mathfrak{B}$. Neither the centralizer, nor the projective representation corresponding to $\mathfrak{M}$, nor the crossed product $\mathfrak{B} = \Delta(\mathfrak{G}, H, \rho)$ is changed if we change the basis (b_s) of $\mathfrak{B}$ to $(b_s\mu_s)$, where the μ_s are non-zero elements of Δ. In particular, if we set $\mu_1 = \rho_{1,1}$ and $\mu_s = 1$ if $s \neq 1$, then the equivalent factor set $\{\rho'_{s,t};\, \bar{u}'\}$ corresponding to the new basis $(b_s\mu_s)$ has the property that $\rho'_{1,1} = 1$, and by an application of (3) (see [6]) it follows that $\rho'_{1,s} = \rho'_{s,1} = 1$. There is no loss of generality in assuming that our original factor set is normalized in this way, and in the rest of the paper, this normalization will be tacitly assumed.

PROPOSITION 1. *The mapping*

$$\tau(\psi, x) = \sum_s b_s\langle \psi, xR_{s^{-1}}\rangle^{\bar{s}}, \tag{7}$$

on $\mathfrak{M}' \times \mathfrak{M} \to \mathfrak{B}$ *is homogeneous, in the sense that the equations*

$$\tau(b\psi, x) = b\tau(\psi, x) \text{ and } \tau(\psi, xb) = \tau(\psi, x)b, \qquad b \in \mathfrak{B}, \tag{8}$$

hold, if and only if the (normalized) factor set of $\mathfrak{B}$ *satisfies the condition* $\rho_{s,s^{-1}} = 1$ *for all* s *in* $\mathfrak{G}$.

Proof. In the proof of this result, we shall use the abbreviation u' for u^{-1}, $u \in \mathfrak{G}$. It is an easy matter to verify that the equations (8) hold if b is an element of Δ. From (7) it follows that $\tau(\psi, x)$ is biadditive, and consequently the homogeneity is equivalent to the equations

$$\tau(b_u\psi, x) = b_u\tau(\psi, x),\ \ \tau(\psi, xb_u) = \tau(\psi, x)b_u, \qquad u \in \mathfrak{G}. \tag{8′}$$

The coefficient of b_t in $\tau(b_u\psi, x)$ is

$$\langle R_u^{*}\psi, xR_{t'}\rangle^{\bar{t}} = \langle \psi, xR_{t'u}\rho_{t',u}\rangle^{\bar{u}'\bar{t}} = \langle \psi, xR_{t'u}\rangle^{\bar{u}'\bar{t}}\rho_{t',u}^{\bar{u}'\bar{t}}.$$

The coefficient of b_t in $b_u\tau(\psi, x)$ is

$$\rho_{u,u't}\langle\psi, xR_{t'u}\rangle^{\overline{u'}\bar{t}} = \rho_{u,u't}\rho_{u',t}\langle\psi, xR_{t'u}\rangle^{\overline{u'}\,\bar{t}}\rho'_{u',t}$$
$$= \rho_{1,t}\rho^{\bar{t}}_{u,u'}\langle\psi, xR_{t'u}\rangle^{\bar{u}'\bar{\rho}_{u,u'}\bar{t}}\rho'_{u',t} = \langle\psi, xR_{t'u}\rangle^{\bar{u}'\bar{t}}\rho^{\bar{t}}_{u,u'}\rho'_{u',t},$$

by (2′) and (3), and the facts that $\rho_{1,t} = 1$, and $\overline{a'} = \bar{u}'\bar{\rho}_{u,u'}$ by (2′). Thus the first equation in (8′) holds if and only if

$$\rho^{\bar{u}'\bar{t}}_{t',u} = \rho^{\bar{t}}_{u,u'}\rho'_{u',t}. \tag{9}$$

The coefficient of b_t in $\tau(\psi, xb_u)$ is $\langle\psi, xR_{ut'}\rho_{u,t'}\rangle^{\bar{t}}$, while the coefficient of b_t in $\tau(\psi, x)b_u$ is

$$\rho_{tu',u}\langle\psi, xR_{ut'}\rangle^{\overline{tu'}\,\bar{u}} = \rho_{tu',u}\rho'_{tu',u}\langle\psi, xR_{ut'}\rangle^{\bar{t}}\rho_{tu',u}$$

by (2). Hence the second equation in (8′) holds if and only if

$$\rho^{\bar{t}}_{u,t'} = \rho_{tu',u}. \tag{10}$$

Setting $t = u$ in (10) we obtain

$$\rho^{\bar{u}}_{u,u'} = \rho_{1,u} = 1,$$

and hence $\rho_{u,u'} = 1$, so that the condition is necessary.

Assume now that $\rho_{u,u'} = 1$ for all u. By (3) we have

$$1 = \rho_{u,1}\rho_{t',t} = \rho_{ut',t}\rho^{\bar{t}}_{u,t'} \tag{11}$$

and

$$1 = \rho_{1,v}\rho_{u'\bar{v},u} = \rho_{u',uv}\rho_{u,v}.$$

Upon substituting $u \to tu'$ and $v \to u$ in the last equation we obtain

$$\rho_{ut',t}\rho_{tu',u} = 1, \tag{12}$$

and by comparing (11) and (12) we obtain (10). The condition implies that $\overline{u'} = \bar{u}'$, and we have

$$\rho^{\overline{u'}}_{t',u}\rho^{\overline{t'}}_{u',t} = \rho_{u't,t'}\rho_{t'u,t'} = 1$$

by (10) and (12), proving (9). This completes the proof.

For an example of a projective representation whose factor set satisfies the condition of Proposition 1, but is not equivalent to one, see (**17**, p. 182).

The pairing $\tau(\psi, x)$ defined in (7) is *non-degenerate in* the sense that $\tau(\mathfrak{M}', x) = 0$ implies $x = 0$, and $\tau(\psi, \mathfrak{M}) = 0$ implies $\psi = 0$. This remark follows from the fact that $\tau(\mathfrak{M}', x) = 0$ implies $\langle\mathfrak{M}', xR_1\rangle = \langle\mathfrak{M}', x\rangle = 0$ since R_1 is the identity operator, and the non-degeneracy of the form $\langle\psi, x\rangle$.

An endomorphism C of $\mathfrak{M}$ is said to belong to the *centralizer* $\mathfrak{C}$ of $\mathfrak{M}$ relative to $\mathfrak{B}$ if $(xb)C = (xC)b$ for all b in $\mathfrak{B}$, x in $\mathfrak{M}$, and if there exists an endomorphism C^* of $\mathfrak{M}'$ such that $\langle C^*\psi, x\rangle = \langle\psi, xC\rangle$ for all x and ψ. An element of $\mathfrak{C}$ is necessarily a l.t. in $\mathfrak{M}$ over Δ, and it follows that C^*, which is uniquely determined, is also a l.t. in $\mathfrak{M}'$ over Δ.

PROPOSITION 2. *An endomorphism C of $\mathfrak{M}$ is an element of $\mathfrak{C}$ if and only if there exists an endomorphism C^{**} on $\mathfrak{M}'$ such that $\tau(C^{**}\psi, x) = \tau(\psi, xC)$ for all x and ψ.*

Proof. If $C \in \mathfrak{C}$ then evidently the transpose C^* of C relative to the form $\langle\psi, x\rangle$ satisfies the equation $\tau(C^*\psi, x) = \tau(\psi, xC)$. Conversely, if C^{**} is given, then upon comparing the coefficients of b_1, we obtain $\langle C^{**}\psi, x\rangle = \langle\psi, xC\rangle$. For all $\xi \in \Delta$,

$$\tau(\psi, (x\xi)C) = \tau(C^{**}\psi, x)\,\xi = \tau(\psi, (xC)\xi),$$

and by the non-degeneracy of the form τ, C is linear. Similarly C^{**} is a l.t., and hence C^{**} is the uniquely determined transpose of C relative to the form $\langle\psi, x\rangle$. Then for all s in $\mathfrak{G}$, comparison of the coefficients of $b_{s^{-1}}$ yields $\langle C^{**}\psi, xR_s\rangle = \langle\psi, xCR_s\rangle$, and hence $\langle\psi, xCR_s\rangle = \langle\psi, xR_sC\rangle$ so that $CR_s = R_sC$ since $\mathfrak{M}'$ and $\mathfrak{M}$ are dual. It follows that C is a $\mathfrak{B}$-endomorphism of $\mathfrak{M}$, and the proof is complete.

We shall call the system $(\mathfrak{M}', \mathfrak{M}, \tau)$ a *pairing* in case τ is bilinear and non-degenerate (τ is bilinear if τ is biadditive and homogeneous relative to right and left multiplication by elements of $\mathfrak{B}$). Necessary and sufficient conditions for the bilinearity of τ are given in Proposition 1. From the bilinearity of τ it follows that

$$\mathfrak{b} = \tau(\mathfrak{M}', \mathfrak{M}) = \{\textstyle\sum \tau(\psi_i, x_i) \mid \psi_i \in \mathfrak{M}', x_i \in \mathfrak{M}\}$$

is a two-sided ideal in $\mathfrak{B}$, which we shall call the *nucleus* of the pairing.

With a pairing $(\mathfrak{M}', \mathfrak{M}, \tau)$, we shall associate a dual pairing $(\psi, u) \to \psi \odot u$ of $\mathfrak{M}' \times \mathfrak{M} \to \mathfrak{C}$, where $\psi \odot u$ is the endomorphism of $\mathfrak{M}$ defined by

$$x(\psi \odot u) = u\tau(\psi, x), \qquad x \in \mathfrak{M}. \tag{13}$$

It is easily verified that if $(\psi \odot u)^*$ is the endomorphism of $\mathfrak{M}'$ defined by

$$(\psi \odot u)^*\phi = \tau(\phi, u)\,\psi, \qquad \phi \in \mathfrak{M}', \tag{14}$$

then $\tau(\phi, x(\psi \odot u)) = \tau((\psi \odot u)^*\phi, x)$, and by Proposition 2, it follows that the mappings $\psi \odot u$ are in $\mathfrak{C}$. The action of $\mathfrak{C}$ upon $\mathfrak{M}$ makes $\mathfrak{M}$ a right $\mathfrak{C}$-module, while $\mathfrak{M}'$ becomes a left $\mathfrak{C}$-module if we set $C\psi = C^*\psi$, where C^* is the transpose of C relative to the forms τ, and $\langle\psi, x\rangle$. It is immediate that the pairing $\psi \odot u$ is bilinear, that is, it is biadditive, and

$$C(\psi \odot u) = (C\psi) \odot u; \qquad (\psi \odot u)C = \psi \odot uC, \qquad C \in \mathfrak{C}.$$

A sufficient condition that $\psi \odot u$ be non-degenerate is that $x \neq 0$, $\psi \neq 0$ imply $x\mathfrak{b} \neq 0$, $\mathfrak{b}\psi \neq 0$, where $\mathfrak{b} = \tau(\mathfrak{M}', \mathfrak{M})$ is the nucleus of the original pairing. Indeed, suppose that $\psi \odot \mathfrak{M} = 0$. Then

$$\mathfrak{M}(\psi \odot \mathfrak{M}) = \mathfrak{M}\tau(\psi, \mathfrak{M}) = 0.$$

Therefore $\tau(\mathfrak{b}\psi, \mathfrak{M}) = 0$, and $\mathfrak{b}\psi = 0$ by the non-degeneracy of τ. Therefore

$\psi = 0$. Similarly $\mathfrak{M}' \odot x = 0$ implies $x = 0$. We have proved the following result.

PROPOSITION 3. *Let $(\mathfrak{M}', \mathfrak{M}, \tau)$ be a pairing. Then $(\psi, u) \to \psi \odot u$ defines a pairing of $\mathfrak{M}' \times \mathfrak{M} \to \mathfrak{C}$ which is bilinear. The pairing $\psi \odot u$ is non-degenerate if $x \neq 0, \psi \neq 0$ imply $x\mathfrak{b} \neq 0$ and $\mathfrak{b}\psi \neq 0$, where $\mathfrak{b}$ is the nucleus of the pairing τ. The set $\mathfrak{c} = \mathfrak{M}' \odot \mathfrak{M}$ consisting of all finite sums $\sum \psi_i \odot u_i$ is a two-sided ideal in $\mathfrak{C}$.*

The mappings $\psi \odot u$ belonging to the nucleus of the pairing defined by (13) can be characterized quite simply if we use the formalism of finite valued l.t. (**8**, Chap. VIII). Every finite valued l.t. X in $\mathfrak{M}$ over Δ which possesses a transpose X^* relative to $\langle\psi, x\rangle$ can be expressed in the form

$$X = \sum \psi_i \times u_i, \qquad \psi_i \in \mathfrak{M}', \qquad u_i \in \mathfrak{M},$$

where $x(\sum \psi_i \times u_i) = \sum u_i\langle\psi_i, x\rangle$, $x \in \mathfrak{M}$. We wish to prove the formula

$$\psi \odot u = \sum_s R_{s^{-1}}(\psi \times u)R_s. \tag{15}$$

We have for all x,

$$x\sum{}_s R_{s^{-1}}(\psi \times u)\, R_s = \sum{}_s u\langle\psi, xR_{s^{-1}}\rangle\, R_s = \sum{}_s uR_s\langle\psi, xR_{s^{-1}}\rangle^{\bar{s}}$$
$$= u\tau(\psi, x) = x(\psi \odot u).$$

Various special cases of the situation considered in this section are of importance. We should like to mention especially the applications to *affine representations* of finite groups (**6**, p. 81), where all $\rho_{s,t} = 1$, and consequently the pairing τ is bilinear in all cases, by Proposition 1; and to ordinary representations of groups, where all $\rho_{s,t} = 1$, all $\bar{s} = 1$, and Δ is a field.

3. A pairing constructed by Weyl. We shall discuss a pairing introduced by Weyl (**15**) which differs from the one we have defined in §2 in that its bilinearity depends upon the existence of an involution in the crossed product $\mathfrak{B}$. We consider an affine representation $s \to U_s$ of a finite group $\mathfrak{G}$ by s.l.t. in a vector space $\mathfrak{M}$ over a field Φ; then all $\rho_{s,t} = 1$, and $s \to \bar{s}$ is a homomorphism. Let $\mathfrak{B} = \Phi(\mathfrak{G}, H, 1)$ be the crossed product constructed as in §2. In this case we have $b_s b_t = b_{st}$, and $\xi b_s = b_s \xi^{\bar{s}}$, $\xi \in \Phi$. Since Φ is commutative it follows that the mapping

$$J: \sum b_s \xi_s \to \sum \xi_s b_{s^{-1}}$$

is an involution in $\mathfrak{B}$. We obtain a representation of $\mathfrak{B}$ by endomorphisms of $\mathfrak{M}$ by setting

$$xU(\textstyle\sum_s b_s \xi_s) = \sum_s (xU_s)\xi_s.$$

Then $\mathfrak{M}$ becomes a left $\mathfrak{B}$-module (and a left vector space over Φ) if we define $bx = xU(b^J)$, $x \in \mathfrak{M}$, $b \in \mathfrak{B}$. The right vector space $\mathfrak{M}^*$ of all linear functions on $\mathfrak{M}$ becomes a right $\mathfrak{B}$-module if we define

$$\psi(\sum b_s \xi_s) = \sum \psi U^*(b_{s^{-1}})\xi_s, \; \psi \in \mathfrak{M}^*,$$

where

$$U^*(b_{s^{-1}})$$

is the transpose of the s.l.t. $U(b_s)$.

We introduce a pairing σ on $\mathfrak{M} \times \mathfrak{M}^* \to \mathfrak{B}$ by means of the following formula:

$$\sigma(x, \psi) = \sum_s b_s \langle xU_s, \psi \rangle, \tag{16}$$

where $\langle x, \psi \rangle$ is the bilinear form on $\mathfrak{M} \times \mathfrak{M}^* \to \Phi$. It is not difficult to verify that σ is bilinear:

$$\begin{aligned} \sigma(x_1 + x_2, \psi) &= \sigma(x_1, \psi) + \sigma(x_2, \psi) \\ \sigma(x, \psi_1 + \psi_2) &= \sigma(x, \psi_1) + \sigma(x, \psi_2), \\ \sigma(bx, \psi) &= b\sigma(x, \psi), \quad \sigma(x, \psi b) = \sigma(x, \psi)b, \qquad b \in \mathfrak{B}, \end{aligned}$$

and that σ is non-degenerate: $\sigma(\mathfrak{M}, \psi) = 0$ implies $\psi = 0$, and $\sigma(x, \mathfrak{M}^*) = 0$ implies $x = 0$.

Let $\mathfrak{C}$ be the ring of $\mathfrak{B}$-endomorphisms of $\mathfrak{M}$. If $C \in \mathfrak{C}$, then C is a l.t. and, if C^* is the transpose of C with respect to the form $\langle x, \psi \rangle$, then

$$\sigma(xC, \psi) = \sigma(x, \psi C^*)$$

for all x and ψ. Conversely if C is a endomorphism of $\mathfrak{M}$, and if there exists an endomorphism C^{**} of $\mathfrak{M}^*$ such that $\sigma(xC, \psi) = \sigma(x, \psi C^{**})$ for all x and ψ, then C^{**} is also the transpose of C with respect to the form $\langle x, \psi \rangle$, and $C \in \mathfrak{C}$.

The endomorphisms $\psi * u$ defined by $x(\psi * u) = \sigma(x, \psi)u$ are elements of $\mathfrak{C}$. If we introduce the action of $\mathfrak{C}$ upon $\mathfrak{M}^*$ by means of the formula $C\psi = \psi C^*$, then $\mathfrak{M}^*$ becomes a left $\mathfrak{C}$-module, and $(\psi, u) \to \psi * u$ defines a bilinear pairing of $\mathfrak{M}^* \times \mathfrak{M} \to \mathfrak{C}$. Finally it is possible to verify, as in §2, that for all ψ and u,

$$\psi * u = \sum_s U_{s^{-1}}(\psi \times u)\, U_s.$$

4. Remarks on the structure and representation theory of crossed products. Let $\mathfrak{B} = \Delta(\mathfrak{G}, H, \rho)$ be a crossed product. We shall prove that there exists a (1–1) order inverting correspondence between the lattices of left and right ideals of $\mathfrak{B}$. Let $r(\mathfrak{S})$ and $l(\mathfrak{S})$ denote the right and left annihilators, respectively, of an arbitrary subset $\mathfrak{S}$ of $\mathfrak{B}$. If $\mathfrak{r}$ and $\mathfrak{l}$ are left and right ideals, respectively, then $r(\mathfrak{l})$ and $l(\mathfrak{r})$ are right and left ideals, respectively.

PROPOSITION 4. *If* $\mathfrak{B} = \Delta(\mathfrak{G}, H, \rho)$, *then the correspondences* $\mathfrak{r} \to l(\mathfrak{r})$ *and* $\mathfrak{l} \to r(\mathfrak{l})$, *where* $\mathfrak{r}$ *and* $\mathfrak{l}$ *are right and left ideals, respectively, are inverses of each other:* $r(l(\mathfrak{r})) = \mathfrak{r}$ *and* $l(r(\mathfrak{l})) = \mathfrak{l}$. *Moreover, every indecomposable right or left ideal direct component of* $\mathfrak{B}$ *contains a unique minimal non-zero subideal.*

Proof. Since $\Delta \subseteq \mathfrak{B}$, $\mathfrak{B}$ is a two-sided vector space over Δ, and the elements $\{b_1, b_2, \ldots\}$ corresponding to the elements of $\mathfrak{G}$ form both a left and right

basis of $\mathfrak{B}$ over Δ. If $b = \sum b_i\xi_i$ is an arbitrary element of $\mathfrak{B}$, then the mapping

$$b \to \lambda(b) = \xi_1$$

is both a left and right Δ-linear function. It is easy to prove that the kernel of λ contains no non-zero left or right ideal of $\mathfrak{B}$ (**12**, p. 658). Therefore the associated bilinear form λ defined by

$$\lambda(b, b') = \lambda(bb'), \qquad b, b' \in \mathfrak{B}. \tag{17}$$

is non-degenerate. From these facts it follows that if $\mathfrak{r}$ and $\mathfrak{l}$ are right and left ideals, respectively, then

$$l(\mathfrak{r}) = \{b \mid b \in \mathfrak{B},\ \lambda(b, \mathfrak{r}) = 0\}, \qquad r(\mathfrak{l}) = \{b \mid b \in \mathfrak{B},\ \lambda(\mathfrak{l}, b) = 0\}.$$

Since $\mathfrak{B}$ is finite dimensional over Δ, a well-known property of dual vector spaces implies the first statement of the theorem.

Now let $e\mathfrak{B} \neq 0$ be an indecomposable right ideal, where e is an idempotent. Since $\mathfrak{B}$ satisfies the minimum condition for left and right ideals, $e\mathfrak{B}$ contains a unique maximal subideal. Moreover $\mathfrak{B}e$ is an indecomposable left ideal which also contains a unique maximal subideal (**1**, Chap. IX). Clearly $l(e\mathfrak{B}) = \mathfrak{B}(1-e)$. Suppose that for some $x \in \mathfrak{B}e$, $\lambda(x, e\mathfrak{B}) = 0$. Then, since $xe\mathfrak{B}$ is a right ideal, we have $xe\mathfrak{B} = 0$, and $x \in \mathfrak{B}(1 - e)$. Therefore $x = 0$, and it follows that the restriction of λ to $\mathfrak{B}e \times e\mathfrak{B}$ is non-degenerate. Because of the order inverting property of the annihilator correspondence, we conclude that both $\mathfrak{B}e$ and $e\mathfrak{B}$ possess unique minimal non-zero subideals.

We remark that $\mathfrak{B}$ is a quasi-Frobenius ring (**10**, p. 8) by Theorem 6 of (**10**).

Now we consider a pairing $(\mathfrak{M}', \mathfrak{M}, \tau)$ of $\mathfrak{M}' \times \mathfrak{M} \to \mathfrak{B}$ (see §2), together with the associated pairing $(\mathfrak{M}', \mathfrak{M}, \odot)$ defined by (13) on $\mathfrak{M}' \times \mathfrak{M}$ to the centralizer $\mathfrak{C}$ of $\mathfrak{M}$ relative to $\mathfrak{B}$. Let $\mathfrak{c} = \mathfrak{M}' \odot \mathfrak{M}$ be the nucleus of the pairing $(\mathfrak{M}', \mathfrak{M}, \odot)$; then $\mathfrak{c}$ is a two-sided ideal in $\mathfrak{C}$. We shall prove that the statement $\mathfrak{c} = \mathfrak{C}$ is equivalent to certain structural properties of $\mathfrak{M}$ viewed as a $\mathfrak{B}$-module. Later, in §8, we shall show how, when $\mathfrak{C} = \mathfrak{c}$, these properties of $\mathfrak{M}$ can be used to prove certain ideal theoretic results concerning the ring $\mathfrak{C}$.

The results we require have been established recently by several authors (**4**; **5**; **9**), and it is unnecessary to include the details here. Let us assume that the (right) dimension of $\mathfrak{M}$ over Δ is finite; then every l.t. X in $\mathfrak{M}$ over Δ has the form $X = \sum \psi_i \times u_i$, for some ψ_i in $\mathfrak{M}'$ and u_i in $\mathfrak{M}$. Our starting point is the observation ((15), §2) that $\mathfrak{c} = \mathfrak{C}$ if and only if there exists a l.t. X in $\mathfrak{M}$ over Δ such that

$$\sum_s R_{s^{-1}} X R_s = 1, \qquad s \in \mathfrak{G}, \tag{18}$$

where 1 is the identity l.t., and R_s is the mapping $x \to xb_s$ in $\mathfrak{M}$.

Now we adopt some terminology due to Cartan and Eilenberg. A (right) $\mathfrak{B}$-module is called *projective*[3] (M_0 in the sense of Gaschutz (**4**; cf. also **5** and

[3]No connection between projective representations and projective modules is implied by this definition.

7) if whenever $\mathfrak{T}$ and $\mathfrak{U}$ are $\mathfrak{B}$-modules such that $\mathfrak{U} \subseteq \mathfrak{T}$ and $\mathfrak{T}/\mathfrak{U} \cong \mathfrak{M}$, then there exists a $\mathfrak{B}$-submodule $\mathfrak{U}^*$ of $\mathfrak{T}$ such that $\mathfrak{T} = \mathfrak{U} \oplus \mathfrak{U}^*$. $\mathfrak{M}$ is called *injective* (M_u in (4, 5, 9)) if whenever $\mathfrak{M}$ is $\mathfrak{B}$-isomorphic to a submodule $\mathfrak{V}$ of $\mathfrak{T}$, then there exists a $\mathfrak{B}$-submodule $\mathfrak{V}^*$ of $\mathfrak{T}$ such that $\mathfrak{T} = \mathfrak{V} \oplus \mathfrak{V}^*$.

PROPOSITION 5. *Let* $(\mathfrak{M}', \mathfrak{M}, \tau)$ *be a pairing of* $\mathfrak{M}' \times \mathfrak{M} \to \mathfrak{B} = \Delta(\mathfrak{G}, H, \rho)$, *and let the (right) dimension of* $\mathfrak{M}$ *over* Δ *be finite. Then the following statements are equivalent.*

(i) $\mathfrak{c}$ *contains the identity l.t.;*
(ii) $\mathfrak{M}$ *is a projective* $\mathfrak{B}$*-module;*
(iii) $\mathfrak{M}$ *is an injective* $\mathfrak{B}$*-module;*
(iv) $\mathfrak{M}$ *is a direct sum of indecomposable* $\mathfrak{B}$*-submodules which are* $\mathfrak{B}$*-isomorphic to right ideal direct components of* $\mathfrak{B}$.

Proof. Theorem 1 of (9) states that (ii) and (iv) are equivalent (see also the remark on p. 107 of (9)). The equivalence of (i), (ii), and (iii) has been proved by Kasch (5, Theorem 12). To verify this statement, the following remarks may be helpful. We should observe first that $\mathfrak{B}$ is a Frobenius extension of Δ with Frobenius homomorphism $b \to \lambda(b)$ (5, p. 462). Then statement (i) is equivalent to the statement that (18) holds for some l.t. X, where we note that $\{b_1, b_s, b_t, \ldots\}$ and $\{b_1, b_{s^{-1}}, b_{t^{-1}}, \ldots\}$ are orthogonal left and right bases of $\mathfrak{B}$ over Δ (5, p. 457) with respect to the bilinear form $\lambda(b, b')$ defined by (17). We now see that Kasch's theorem is indeed applicable to our situation.

Remark 1. If it is not assumed that the dimension of $\mathfrak{M}$ over Δ is finite, then not every l.t. X in $\mathfrak{M}$ over Δ has the form $\sum \psi_i x u_i$. The following implications remain valid: (i) $\to$ (18) $\to$ [(ii) and (iii)] $\to$ (iv).

Remark 2. Assume (i); then from (18) we obtain

$$\sum L_{s^{-1}} X^* L_s = 1,$$

where 1 is now the identity mapping on $\mathfrak{M}'$, X^* is a l.t. on $\mathfrak{M}'$, and L_s is the mapping $\psi \to b_s \psi = R_s^* \psi$ in $\mathfrak{M}'$. Therefore we have the implications (i) $\to$ (ii)$'$ $\to$ (iv)$'$, where (ii)$'$ and (iv)$'$ are obtained from (ii) and (iv) by replacing $\mathfrak{M}$ by $\mathfrak{M}'$, and "right" by "left" in (iv).

Remark 3. It follows from the considerations of §3 that a result analogous to Proposition 5 can be established for the pairing σ of $\mathfrak{M} \times \mathfrak{M}^* \to \mathfrak{B}$ which was constructed in §3. We shall not include the details of this discussion.

5. Abstract theory of regular pairings. Let $\mathfrak{B}$ be an arbitrary ring with identity element 1, and let $\mathfrak{B}$ admit a set of Ω of (left) operators. We shall assume that 1 acts as the identity operator on all $\mathfrak{B}$-modules which we shall consider. Let $\mathfrak{M}'$ and $\mathfrak{M}$ be left and right $\mathfrak{B}$-Ω-modules, which are paired to $\mathfrak{B}$ by a function $\tau(\psi, x)$. We assume that τ is bilinear, relative to both $\mathfrak{B}$ and Ω, in the sense that the equations

$$\tau(\psi_1 + \psi_2, x) = \tau(\psi_1, x) + \tau(\psi_2, x), \ \tau(\psi, x_1 + x_2) = \tau(\psi, x_1) + \tau(\psi, x_2)$$
$$\tau(b\psi, x) = b\tau(\psi, x), \quad \tau(\psi, xb) = \tau(\psi, x)b,$$
$$\tau(\alpha\psi, x) = \alpha\tau(\psi, x), \quad \tau(\psi, \alpha x) = \alpha\tau(\psi, x)$$

hold for all x in $\mathfrak{M}$, ψ in $\mathfrak{M}'$, b in $\mathfrak{B}$, and α in Ω. Our second assumption is that τ is non-degenerate. If these conditions are satisfied, then we shall call the system $(\mathfrak{M}', \mathfrak{M}, \tau)$ an *(abstract) pairing*. The *nucleus* $\mathfrak{b} = \tau(\mathfrak{M}', \mathfrak{M})$ of the pairing is a two-sided ideal in $\mathfrak{B}$.

We let $\mathfrak{E}$ be the set of all $\mathfrak{B}$-Ω-endomorphisms of $\mathfrak{M}$. If $\bar{\alpha}$ denotes the endomorphism $x \to \alpha x$ of $\mathfrak{M}$ determined by an element of Ω, then $\bar{\alpha}E \in \mathfrak{E}$ for every E in $\mathfrak{E}$, and $\bar{\alpha}E = E\bar{\alpha}$, so that if we define $\alpha E = \bar{\alpha}E$, then $\mathfrak{E}$ becomes an Ω-ring.

The endomorphisms $\psi \odot u$ defined by (13) are elements of $\mathfrak{E}$, and possess transposes relative to the form τ. Let $\mathfrak{c}$ be the subgroup of $\mathfrak{E}$ consisting of all finite sums of the $\psi \odot u$. If the action of $\mathfrak{c}$ upon $\mathfrak{M}'$ is defined by the formula $E\psi = E^*\psi$, for E in $\mathfrak{c}$, then it follows that $(\psi, u) \to \psi \odot u$ is a $\mathfrak{c}$-Ω-bilinear mapping of $\mathfrak{M}' \times \mathfrak{M} \to \mathfrak{c}$. We shall denote this pairing by $(\mathfrak{M}', \mathfrak{M}, \odot)$, and observe that the nucleus $\mathfrak{c}$ is an Ω-subring of $\mathfrak{E}$. If $E \in \mathfrak{E}$, then $(\psi \odot u)E = \psi \odot uE$, and hence $\mathfrak{c}$ is a right Ω-ideal in $\mathfrak{E}$. We shall denote by $\mathfrak{C}$ an arbitrary Ω-subring of $\mathfrak{E}$ such that

$$\mathfrak{c} \subseteq \mathfrak{C} \subseteq \mathfrak{E}. \tag{19}$$

Then $\mathfrak{C}$ will be called a *centralizer* of $\mathfrak{M}$ relative to $\mathfrak{B}$, and will remain fixed throughout the discussion. Our aim is to establish relationships between the nuclei $\mathfrak{b}$ and $\mathfrak{c}$ of the rings $\mathfrak{B}$ and $\mathfrak{C}$, and the properties of $\mathfrak{M}$ and $\mathfrak{M}'$ as $\mathfrak{B}$ and $\mathfrak{C}$-modules.

In order to discuss the connection between the ring $\mathfrak{B}$ and the structure of $\mathfrak{M}$ (or $\mathfrak{M}'$) as a $\mathfrak{C}$-module, we shall assume that the pairing τ is *regular* in the sense that $\mathfrak{b}$ contains an element $e_0 = \sum\tau(\psi^*_i, x^*_i)$ such that $be_0 = e_0b = b$ for all $b \in \mathfrak{b}$. By the non-degeneracy of τ it follows that $xe_0 = x$ and $e_0\psi = \psi$ for all $x \in \mathfrak{M}$ and $\psi \in \mathfrak{M}'$.

It is always possible to construct a regular pairing from an arbitrary one. Let e_0 be any central idempotent contained in the nucleus $\mathfrak{b}$ of a pairing $(\mathfrak{M}', \mathfrak{M}, \tau)$, or let $e_0 = 0$ if $\mathfrak{b}$ contains no central idempotent. Then

$$\mathfrak{M} = \mathfrak{M}e_0 \oplus \mathfrak{M}(1 - e_0), \qquad \mathfrak{M}' = e_0\mathfrak{M}' \oplus (1 - e_0)\mathfrak{M}',$$

where the direct summands are invariant relative to both $\mathfrak{B}$ and $\mathfrak{c}$. We define a new pairing τ_0 of $e_0\mathfrak{M}' \times \mathfrak{M}e_0 \to \mathfrak{B}$ by setting

$$\tau_0(e_0\psi, xe_0) = \tau(e_0\psi, xe_0)$$

for all ψ and x and we shall prove that τ_0 is a regular pairing. The nucleus $\mathfrak{b}_0$ of τ_0 contains e_0, for if $e_0 = \sum \tau(\psi_i, x_i)$, then $e_0 = \sum\tau(e_0\psi_i, x_ie_0)$. Obviously

$$e_0b = be_0 = b, \qquad b \in \mathfrak{b}_0.$$

The bilinearity of τ_0 is evident. It remains to prove that τ_0 is non-degenerate.

Suppose $\tau_0(e_0\psi, xe_0) = 0$ for all $e_0\psi \in e_0\mathfrak{M}'$. If ψ is arbitrary in $\mathfrak{M}'$, we write

$$\psi = e_0\psi + (1 - e_0)\psi$$

and obtain

$$\begin{aligned}\tau(\psi, xe_0) &= \tau_0(e_0\psi, xe_0) + \tau((1 - e_0)\psi, xe_0)\\ &= \tau((1 - e_0)\psi, xe_0)e_0 = e_0\tau((1 - e_0)\psi, xe_0) = 0,\end{aligned}$$

so that $xe_0 = 0$ by the non-degeneracy of τ. Similarly $\tau_0(e_0\psi, \mathfrak{M}e_0) = 0$ implies $e_0\psi = 0$.

We return now to our assumption that the pairing is regular. If $\mathfrak{S}$ is any subset of $\mathfrak{B}$, we shall write $\mathfrak{S}_r$ (resp. $\mathfrak{S}_l$) for the set of endomorphisms R_s: $x \to xs$ (resp. L_s: $\psi \to s\psi$) of $\mathfrak{M}$ (resp. $\mathfrak{M}'$) determined by the elements of $\mathfrak{S}$. We are in a position to prove the following result:

THEOREM 1. *Let $\overline{\mathfrak{B}}$ be the set of all $\mathfrak{C}$-endomorphisms of $\mathfrak{M}$. Then $\mathfrak{b}_r = \mathfrak{B}_r = \overline{\mathfrak{B}}$.*

Proof. Obviously $\mathfrak{b}_r \subseteq \mathfrak{B}_r \subseteq \overline{\mathfrak{B}}$. Conversely let $B \in \overline{\mathfrak{B}}$; then $B(\psi \odot u) = (\psi \odot u)B$ for all ψ and u. Consequently

$$u\tau(\psi, xB) = (u\tau(\psi, x))B$$

for all x, ψ, and u. Let $b = \sum\tau(\psi^*{}_i, x^*{}_iB)$; then for all $u \in \mathfrak{M}$ we have

$$ub = \sum u\tau(\psi^*{}_i, x^*{}_iB) = (u\sum\tau(\psi^*{}_i, x^*{}_i))B = (ue_0)B = uB,$$

and $R_b = B$. This completes the proof.

If $\mathfrak{N}$ is a $\mathfrak{C}$-Ω-submodule of $\mathfrak{M}$, then

$$\tau(\mathfrak{M}', \mathfrak{N}) = \{\sum \tau(\psi_i, x_i) \mid \psi_i \in \mathfrak{M}', x_i \in \mathfrak{N}\}$$

is a left Ω-ideal contained in $\mathfrak{b}$. If $\mathfrak{l}$ is a left Ω-ideal in $\mathfrak{B}$, then $\mathfrak{M}\mathfrak{l}$ is a $\mathfrak{C}$-Ω-submodule of $\mathfrak{M}$. We have, for all $\mathfrak{N}$ and $\mathfrak{l}$,

$$\mathfrak{M}\tau(\mathfrak{M}', \mathfrak{N}) \subseteq \mathfrak{N}; \qquad \tau(\mathfrak{M}', \mathfrak{M}\mathfrak{l}) \subseteq \mathfrak{l}: \tag{20}$$

the first since $\mathfrak{M}\tau(\mathfrak{M}', \mathfrak{N}) \subseteq \mathfrak{N}(\mathfrak{M}' \odot \mathfrak{M}) \subseteq \mathfrak{N}\mathfrak{C} \subseteq \mathfrak{N}$ by (19) and the fact that $\mathfrak{N}$ is a $\mathfrak{C}$-submodule[4] of $\mathfrak{M}$; the second, obvious. For later use we observe also that

$$\tau(\mathfrak{M}', \sum \mathfrak{N}_\nu) = \sum \tau(\mathfrak{M}', \mathfrak{N}_\nu), \quad \mathfrak{M}(\sum \mathfrak{l}_\nu) = \sum (\mathfrak{M}\mathfrak{l}_\nu), \tag{21}$$

and

$$\tau(\mathfrak{M}', \mathfrak{N}\mathfrak{l}) = \tau(\mathfrak{M}', \mathfrak{N})\mathfrak{l}, \qquad \mathfrak{M}(\mathfrak{l}_1\mathfrak{l}_2) = (\mathfrak{M}\mathfrak{l}_1)\mathfrak{l}_2. \tag{22}$$

LEMMA 1. *Let $\mathfrak{N}$ be a $\mathfrak{C}$-direct summand of $\mathfrak{M}$. Then there exists an idempotent $e \in \mathfrak{B}$ such that $\tau(\mathfrak{M}', \mathfrak{N}) = \mathfrak{B}e$.*

---

[4]For the rest of §5, 6, and 7, we shall omit explicit reference to the set Ω. Thus by submodule, ideal, etc. we shall mean Ω-submodule, Ω-ideal, etc.

Proof. Let E be a projection of $\mathfrak{M}$ upon $\mathfrak{N}$ such that $E \in \overline{\mathfrak{B}}$. By Theorem 1, $E = R_e$, where

$$e = \sum \tau(\psi^*_i, x^*_i E) \in \tau(\mathfrak{M}', \mathfrak{N}).$$

If $b = \sum \tau(\psi_i, x_i)$ is an arbitrary element of $\tau(\mathfrak{M}', \mathfrak{N})$, then $be = b$ since the restriction of E to $\mathfrak{N}$ is the identity mapping. Therefore $\tau(\mathfrak{M}', \mathfrak{N}) = \mathfrak{B}e$.

LEMMA 2. *Let $\mathfrak{N}$ be a $\mathfrak{C}$-submodule such that $\tau(\mathfrak{M}', \mathfrak{N}) = \mathfrak{B}e$, where e is an idempotent in $\mathfrak{B}$. Then $\mathfrak{M}\tau(\mathfrak{M}', \mathfrak{N}) = \mathfrak{N}$.*

Proof. By the non-degeneracy of τ we have $x = xe \in \mathfrak{M}\tau(\mathfrak{M}', \mathfrak{N})$ for all $x \in \mathfrak{N}$, and together with (20), this proves the Lemma.

LEMMA 3. *Let $\mathfrak{l} = \mathfrak{B}e$, where $e^2 = e \in \mathfrak{b}$. Then $\mathfrak{M}\mathfrak{l} = \mathfrak{M}e$ is a $\mathfrak{C}$-direct summand of $\mathfrak{M}$, and $\tau(\mathfrak{M}', \mathfrak{M}\mathfrak{l}) = \mathfrak{l}$.*

Proof. We have $\mathfrak{M}\mathfrak{l} = \mathfrak{M}\mathfrak{B}e = \mathfrak{M}e$, and $\mathfrak{M} = \mathfrak{M}e \oplus \mathfrak{M}(1 - e)$, proving the first statement. For the second, $b \in \mathfrak{l}$, $b = \sum \tau(\psi_i, x_i)$, implies

$$b = be = \sum \tau(\psi_i, x_i e) \in \tau(\mathfrak{M}', \mathfrak{M}\mathfrak{l}),$$

and by (20) we infer that $\mathfrak{l} = \tau(\mathfrak{M}', \mathfrak{M}\mathfrak{l})$.

THEOREM 2. *Let $(\mathfrak{M}', \mathfrak{M}, \tau)$ be a regular pairing with nucleus $\mathfrak{b}$. The mappings $\mathfrak{l} \to \mathfrak{M}\mathfrak{l}$ and $\mathfrak{N} \to \tau(\mathfrak{M}', \mathfrak{N})$ between the set of left ideal direct components of $\mathfrak{b}$ and the $\mathfrak{C}$-direct summands of $\mathfrak{M}$ are inverses of each other. The mapping $\mathfrak{l} \to \mathfrak{M}\mathfrak{l}$ preserves sums of arbitrary ideals, and intersections of left ideal direct components of $\mathfrak{B}$. Two left ideal direct components $\mathfrak{l}_1$ and $\mathfrak{l}_2$ of $\mathfrak{b}$ are $\mathfrak{B}$-isomorphic if and only if $\mathfrak{M}\mathfrak{l}_1$ and $\mathfrak{M}\mathfrak{l}_2$ are $\mathfrak{C}$-isomorphic.*

Proof. The first statement follows from Lemmas 1–3. By (21) the mapping $\mathfrak{l} \to \mathfrak{M}\mathfrak{l}$ preserves sums. The statement concerning intersections is an immediate consequence of the fact to be proved next, that if $\mathfrak{l}$ is a left ideal direct component of $\mathfrak{b}$ then

$$\mathfrak{M}\mathfrak{l} = \{x \mid \tau(\mathfrak{M}', x) \subseteq \mathfrak{l}\}.$$

Let $\mathfrak{l} = \mathfrak{B}e$, where $e^2 = e \in \mathfrak{b}$. Then $\tau(\mathfrak{M}', x) \subseteq \mathfrak{l}$ implies $\tau(\psi, xe) = \tau(\psi, x)$ for all $\psi \in \mathfrak{M}'$, and by the non-degeneracy of τ, $x = xe \in \mathfrak{M}\mathfrak{l}$. Conversely $x \in \mathfrak{M}\mathfrak{l}$ implies $xe = x$, and

$$\tau(\mathfrak{M}', x) = \tau(\mathfrak{M}', x)e \subseteq \mathfrak{l}.$$

Let $\mathfrak{B}e_1$ and $\mathfrak{B}e_2$ be $\mathfrak{B}$-isomorphic; then there exist elements a and b such that

$$\mathfrak{B}e_1 a = \mathfrak{B}e_2, \quad \mathfrak{B}e_2 b = \mathfrak{B}e_1, \quad cab = c, \qquad c \in \mathfrak{B}e_1,$$

$dba = d$ for all $d \in \mathfrak{B}e_2$. One verifies easily that $xe_1 \to xe_1 a$ and $xe_2 \to xe_2 b$ are $\mathfrak{C}$-homomorphisms between $\mathfrak{M}e_1$ and $\mathfrak{M}e_2$ which are inverses of each other, and consequently $\mathfrak{M}e_1$ and $\mathfrak{M}e_2$ are $\mathfrak{C}$-isomorphic.

Conversely let $x \to x^h$ be a $\mathfrak{C}$-isomorphism of $\mathfrak{N}_1$ onto $\mathfrak{N}_2$. Define

$$\bar{h}: \sum \tau(\psi_i, x_i) \to \sum \tau(\psi_i, x_i^h)$$

of $\tau(\mathfrak{M}', \mathfrak{N}_1)$ into $\tau(\mathfrak{M}', \mathfrak{N}_2)$. In order to prove that $\tau(\mathfrak{M}', \mathfrak{N}_1)$ and $\tau(\mathfrak{M}', \mathfrak{N}_2)$ are $\mathfrak{B}$-isomorphic, it is clearly sufficient to prove that $\bar{h}$ is a $\mathfrak{B}$-homomorphism onto. If

$$\sum \tau(\psi_i, x_i) = 0, \quad x_i \in \mathfrak{N}_1,$$

then $0 = \mathfrak{M}(\sum \tau(\psi_i, x_i)) = \sum x_i(\psi_i \odot \mathfrak{M})$, and since h is a $\mathfrak{C}$-isomorphism,

$$\sum x_i^h(\psi_i \odot \mathfrak{M}) = \sum \mathfrak{M}\tau(\psi_i, x_i^h) = 0.$$

Since $e_0 = \sum \tau(\psi^*_i, x^*_i)$ is a left identity element in $\mathfrak{b}$, we have

$$\sum \tau(\psi_i, x_i^h) = \sum e_0 \tau(\psi_i, x_i^h) = 0.$$

Thus $\bar{h}$ is single valued. The fact that it is onto, and is a $\mathfrak{B}$-homomorphism can be checked in a similar way using the properties of τ. This completes the proof.

COROLLARY. *A left ideal direct component $\mathfrak{l}$ of $\mathfrak{b}$ is indecomposable if and only if $\mathfrak{M}\mathfrak{l}$ is an indecomposable direct summand of $\mathfrak{M}$.*

Proof. Let $\mathfrak{l}$ be a decomposable direct component of $\mathfrak{B}$: $\mathfrak{l} = \mathfrak{l}_1 \oplus \mathfrak{l}_2$. Then by the theorem $\mathfrak{M}\mathfrak{l} = \mathfrak{M}\mathfrak{l}_1 \oplus \mathfrak{M}\mathfrak{l}_2$, where neither component is zero. The converse is proved similarly.

Let us denote by $\mathfrak{c}^*$ the set of transposes relative to τ of the elements of $\mathfrak{c}$, and write $\mathfrak{B}_l$ and $\mathfrak{b}_l$, respectively, for the sets of endomorphisms $\psi \rightarrow b\psi$ determined in $\mathfrak{M}'$ by the elements of $\mathfrak{B}$ and $\mathfrak{b}$. Let $\mathfrak{E}^*$ be the set of all $\mathfrak{B}$-Ω-endomorphisms of $\mathfrak{M}'$, and let $\mathfrak{C}^*$ be an arbitrary ring of Ω-endomorphisms of $\mathfrak{M}'$ such that $\mathfrak{c}^* \subseteq \mathfrak{C}^* \subseteq \mathfrak{E}^*$. We shall write $\mathfrak{B}'$ for the set of $\mathfrak{C}^*$-endomorphisms of $\mathfrak{M}'$. Then we may state the following duals to Theorems 1 and 2.

THEOREM 1′. *Let $(\mathfrak{M}', \mathfrak{M}, \tau)$ be a regular pairing with nucleus $\mathfrak{b}$, and let $\mathfrak{C}^*$ be an Ω-subring of $\mathfrak{E}^*$ containing $\mathfrak{c}^*$. Then $\mathfrak{b}_l = \mathfrak{B}_l = \mathfrak{B}'$.*

THEOREM 2′. *Let $(\mathfrak{M}', \mathfrak{M}, \tau)$ be a regular pairing with nucleus $\mathfrak{b}$. The mappings $\mathfrak{r} \rightarrow \mathfrak{r}\mathfrak{M}'$, $\mathfrak{N}' \rightarrow \tau(\mathfrak{N}', \mathfrak{M})$ between the sets of right ideal direct components of $\mathfrak{b}$ and the $\mathfrak{C}^*$-direct summands of $\mathfrak{M}'$ are inverses of each other, and possess the properties stated in Theorem 2.*

THEOREM 3 (Weyl).[5] *Let $(\mathfrak{M}', \mathfrak{M}, \tau)$ be a pairing of $\mathfrak{M}' \times \mathfrak{M} \rightarrow \mathfrak{B}$, where $\mathfrak{B}$ is a semi-simple Ω-ring satisfying the minimum condition for left ideals. Then the pairing is regular. The mappings $\mathfrak{l} \rightarrow \mathfrak{M}\mathfrak{l}$ and $\mathfrak{N} \rightarrow \tau(\mathfrak{M}', \mathfrak{N})$ are inverses of each other, and establish a (1–1) inclusion preserving correspondence between the set of all left ideals of $\mathfrak{B}$ which are contained in the nucleus $\mathfrak{b}$, and the set of all $\mathfrak{C}$-submodules of $\mathfrak{M}$. If*

$$\mathfrak{l}_1 \leftrightarrow \mathfrak{N}_1 = \mathfrak{M}\mathfrak{l}_1, \quad \mathfrak{l}_2 \leftrightarrow \mathfrak{N}_2 = \mathfrak{M}\mathfrak{l}_2,$$

[5] This result, and Theorem 2 in its essentials, have been proved by Weyl for pairings of the type considered in §3 (**16**, Chap. 5; **15**; **17**, Chap. 3).

then

$$\mathfrak{l}_1 + \mathfrak{l}_2 \leftrightarrow \mathfrak{N}_1 + \mathfrak{N}_2, \qquad \mathfrak{l}_1 \cap \mathfrak{l}_2 \leftrightarrow \mathfrak{N}_1 \cap \mathfrak{N}_2,$$

and $\mathfrak{l}_1$ *and* $\mathfrak{l}_2$ *are* $\mathfrak{B}$*-isomorphic if and only if* $\mathfrak{N}_1$ *and* $\mathfrak{N}_2$ *are* $\mathfrak{C}$*-isomorphic.*

Proof. The structure theory of semi-simple rings implies that the pairing is regular, and that every left ideal in $\mathfrak{b}$ is a direct component of $\mathfrak{b}$. By Theorem 2, $\mathfrak{l} \to \mathfrak{M}\mathfrak{l}$ is a (1–1) inclusion preserving correspondence between the set of all left ideal direct components of $\mathfrak{b}$ and the set of all $\mathfrak{C}$-submodules of $\mathfrak{M}$. By a principle of lattice theory, the mapping preserves the lattice operations. That it preserves isomorphism relations has been proved in Theorem 2.

Example. Let $b \to U(b)$ be an ordinary representation of the group algebra $\mathfrak{B}$ of a finite group $\mathfrak{G}$ by l.t. in a finite dimensional vector space $\mathfrak{M}$ over a field, and let $\mathfrak{C}$ be the set of all l.t. commuting with the l.t. $U(b)$, $b \in \mathfrak{B}$. Let $\mathfrak{b}_r$ and $\mathfrak{b}_s$ be the nuclei of the pairings constructed in **2** and **3** respectively. Finally let us assume that both pairings are regular. Then by Theorems 2 and 2′, a left ideal $\mathfrak{B}e$ of $\mathfrak{b}_r$ is matched against the $\mathfrak{C}$-submodule $\mathfrak{M}U(e)$ of $\mathfrak{M}$, while a right ideal $f\mathfrak{B}$ of $\mathfrak{b}_s$ generated by an idempotent f is matched against the $\mathfrak{C}$-submodule $\mathfrak{M}U(f^J)$. We remark finally that $\mathfrak{b}_s = \mathfrak{b}_r{}^J$.

6. Maximal submodules of indecomposable $\mathfrak{C}$-direct summands. We adhere to the assumptions and notation of §5, and make the additional assumption that $\mathfrak{B}$ satisfies the minimum condition for left ideals, and hence also the maximum condition, since $\mathfrak{B}$ has an identity element. Let $\mathfrak{R}$ be the radical of $\mathfrak{B}$; then every indecomposable left ideal direct component $\mathfrak{B}e$ of $\mathfrak{B}$ has a unique maximal subideal $\mathfrak{R}e$. Every proper subideal of $\mathfrak{B}e$ is nilpotent, and $\mathfrak{B}e$ and $\mathfrak{B}e'$ are $\mathfrak{B}$-isomorphic if and only if $\mathfrak{B}e/\mathfrak{R}e$ and $\mathfrak{B}e'/\mathfrak{R}e'$ are $\mathfrak{B}$-isomorphic (**1**, Chap. IX).

LEMMA 4. *Let* $\mathfrak{N} = \mathfrak{M}e$ *be an indecomposable* $\mathfrak{C}$*-direct summand of* $\mathfrak{M}$. *Then* $\mathfrak{N}$ *has a unique maximal* $\mathfrak{C}$*-submodule* $\mathfrak{S}$, *and*

$$\tau(\mathfrak{M}', \mathfrak{S}) \subseteq \mathfrak{R}e, \quad \mathfrak{M}(\mathfrak{R}e) \subseteq \mathfrak{S}. \tag{23}$$

Proof. By the Corollary to Theorem 2, $\mathfrak{B}e = \tau(\mathfrak{M}', \mathfrak{N})$ is an indecomposable left ideal. Let $\mathfrak{S} = \sum \mathfrak{N}_\nu$, where $\{\mathfrak{N}_\nu\}$ is the set of all proper $\mathfrak{C}$-submodules of $\mathfrak{N}$. By (21) and the fact that $\mathfrak{B}$ satisfies the maximum condition for left ideals, we have

$$\tau(\mathfrak{M}', \mathfrak{S}) = \sum \tau(\mathfrak{M}', \mathfrak{N}_\nu),$$

which in turn can be expressed as a finite sum

$$\textstyle\sum_{i=1}^{m} \tau(\mathfrak{M}', \mathfrak{N}_{\nu_i}).$$

No $\tau(\mathfrak{M}', \mathfrak{N}_\nu) = \mathfrak{B}e$, otherwise, by Lemma 2,

$$\mathfrak{N}_\nu = \mathfrak{M}\tau(\mathfrak{M}', \mathfrak{N}_\nu) = \mathfrak{M}e = \mathfrak{N}.$$

Hence each $\tau(\mathfrak{M}', \mathfrak{N}_\nu)$ is nilpotent, and since the sum is finite, $\tau(\mathfrak{M}', \mathfrak{S})$ is nilpotent. This proves (i) $\mathfrak{S} \neq \mathfrak{N}$ (for if $\mathfrak{S} = \mathfrak{N}$ then $\tau(\mathfrak{M}', \mathfrak{S})$ contains an idempotent $\neq 0$) and (ii), $\tau(\mathfrak{M}', \mathfrak{S}) \subseteq \mathfrak{N}e$. For the other inclusion of (23) it is sufficient to prove that $\mathfrak{M}(\mathfrak{N}e) \neq \mathfrak{N}$. If, however, $\mathfrak{M}(\mathfrak{N}e) = \mathfrak{N}$, then by Lemma 3 and (22) we have

$$\mathfrak{B}e = \tau(\mathfrak{M}', \mathfrak{N}) = \tau(\mathfrak{M}', \mathfrak{M})\mathfrak{N}e \subseteq \mathfrak{N},$$

contrary to our assumption that $e^2 = e \neq 0$. This completes the proof.

THEOREM 4. *Let $\mathfrak{N}_1$ and $\mathfrak{N}_2$ be indecomposable $\mathfrak{C}$-direct summands of $\mathfrak{M}$ with maximal $\mathfrak{C}$-submodules $\mathfrak{S}_1$ and $\mathfrak{S}_2$. Then $\mathfrak{N}_1/\mathfrak{S}_1$ and $\mathfrak{N}_2/\mathfrak{S}_2$ are $\mathfrak{C}$-isomorphic if and only if $\mathfrak{N}_1$ and $\mathfrak{N}_2$ are $\mathfrak{C}$-isomorphic.*

Proof. We prove the result by throwing the argument back to the known results concerning the ideals in $\mathfrak{B}$. Using Lemma 4, it is easy to prove that the $\mathfrak{C}$-isomorphism of $\mathfrak{N}_1$ onto $\mathfrak{N}_2$ induces a $\mathfrak{C}$-isomorphism of $\mathfrak{N}_1/\mathfrak{S}_1$ onto $\mathfrak{N}_2/\mathfrak{S}_2$. For the proof of the converse it is enough to show, by Theorem 2, that $\mathfrak{B}e_1 = \tau(\mathfrak{M}', \mathfrak{N}_1)$ and $\mathfrak{B}e_2 = \tau(\mathfrak{M}', \mathfrak{N}_2)$ are $\mathfrak{B}$-isomorphic. This we prove by showing that $\mathfrak{B}e_1/\mathfrak{N}e_1$ and $\mathfrak{B}e_2/\mathfrak{N}e_2$ are $\mathfrak{B}$-isomorphic, assuming that $\mathfrak{N}_1/\mathfrak{S}_1$ and $\mathfrak{N}_2/\mathfrak{S}_2$ are $\mathfrak{C}$-isomorphic.

Let ζ be a $\mathfrak{C}$-isomorphism of $\mathfrak{N}_1/\mathfrak{S}_1$ onto $\mathfrak{N}_2/\mathfrak{S}_2$, and let $\theta = \zeta^{-1}$. In both $\mathfrak{N}_1$ and $\mathfrak{N}_2$ select a fixed system of representatives of the cosets in $\mathfrak{N}_1/\mathfrak{S}_1$ and $\mathfrak{N}_2/\mathfrak{S}_2$ respectively, and for each $x_1 \in \mathfrak{N}_1$, let $x_1\bar{\zeta}$ be the representative of the coset $(x_1 + \mathfrak{S}_1)\zeta$; that is

$$x_1\bar{\zeta} + \mathfrak{S}_2 = (x_1 + \mathfrak{S}_1)\zeta.$$

Similarly we define a map $\bar{\theta}$ of $\mathfrak{N}_2$ into $\mathfrak{N}_1$. We have

$$x_1 \equiv x_1\bar{\zeta}\bar{\theta} \pmod{\mathfrak{S}_1}, \quad x_2 \equiv x_2\bar{\theta}\bar{\zeta} \pmod{\mathfrak{S}_2}, \tag{24}$$

for all $x_1 \in \mathfrak{N}_1$, $x_2 \in \mathfrak{N}_2$.

Now define a mapping μ of $\tau(\mathfrak{M}', \mathfrak{N}_1)$ into $\tau(\mathfrak{M}', \mathfrak{N}_2)$, namely

$$\mu: \sum \tau(\psi_i, x_{1i}) \to \sum \tau(\psi_i, x_{1i}\bar{\zeta}),$$

where $x_{1i} \in \mathfrak{N}_1$, $\psi_i \in \mathfrak{M}'$ for all i. We contend that the induced mapping

$$\bar{\mu}: \sum \tau(\psi_i, x_{1i}) + \mathfrak{N}e_1 \to \sum \tau(\psi_i, x_{1i}\bar{\zeta}) + \mathfrak{N}e_2$$

is a $\mathfrak{B}$-isomorphism of $\mathfrak{B}e_1/\mathfrak{N}e_1$ onto $\mathfrak{B}e_2/\mathfrak{N}e_2$.

First we prove that $\bar{\mu}$ is single valued. Let $a = \sum \tau(\psi_i, x_{1i}) \in \mathfrak{N}e_1$; then $\mathfrak{M}a \subseteq \mathfrak{M}(\mathfrak{N}e_1) \subseteq \mathfrak{S}_1$ by (23). Thus for all $u \in \mathfrak{M}$,

$$ua + \mathfrak{S}_1 = \sum x_{1i}(\psi_i \odot u) + \mathfrak{S}_1 = 0.$$

Applying ζ we have $\sum (x_{1i} + \mathfrak{S}_1)\zeta(\psi_i \odot u) = 0$. Then

$$\sum x_{1i}\bar{\zeta}(\psi_i \odot u) \in \mathfrak{S}_2, \quad \sum \mathfrak{M}\tau(\psi_i, x_{1i}\bar{\zeta}) \subseteq \mathfrak{S}_2.$$

If $e_0 = \sum \tau(\psi^*_i, x^*_i)$ is the identity element in $\mathfrak{b}$, then $a = e_0 a$ implies

$$a\mu = e_0(a\mu) \in \tau(\mathfrak{M}', \mathfrak{S}_2) \subseteq \mathfrak{R}e_2$$

by (23), and $\bar{\mu}$ is single valued. That $\bar{\mu}$ is a $\mathfrak{B}$-homomorphism follows from the bilinearity of τ, and the ontoness from (24) and (23). To prove that $\bar{\sigma}$ is (1–1), let

$$\sum \tau(\psi_i, x_{1i}\xi) \in \mathfrak{R}e_2, \qquad \psi_i \in \mathfrak{M}', \qquad x_{1i} \in \mathfrak{R}_1.$$

Then by (23), $\sum \mathfrak{M}\tau(\psi_i, x_{1i}\xi) \subseteq \mathfrak{S}_2$; as in the first part of the proof we now verify that $\sum \tau(\psi_i, x_{1i}\xi\theta) \in \mathfrak{R}e_1$, and that

$$\sum \tau(\psi_i, x_{1i}) - \sum \tau(\psi_i, x_{1i}\xi\theta) = \sum \tau(\psi_i, x_{1i} - x_{1i}\xi\theta) \in \tau(\mathfrak{M}', \mathfrak{S}_1) \subseteq \mathfrak{R}e_1$$

by (23). Thus $\sum \tau(\psi_i, x_{1i}) \in \mathfrak{R}e_1$, and we have proved that $\bar{\mu}$ is (1–1). This completes the proof of the theorem.

7. The structure of $\mathfrak{c} = \mathfrak{M}' \odot \mathfrak{M}$. Let $(\mathfrak{M}', \mathfrak{M}, \tau)$ be an abstract pairing. We shall assume that the nucleus $\mathfrak{c} = \mathfrak{M}' \odot \mathfrak{M}$ of the associated pairing $(\mathfrak{M}', \mathfrak{M}, \odot)$ contains the identity mapping on $\mathfrak{M}$, and that the function $\psi \odot u$ is non-degenerate. Since $\mathfrak{c}$ is a right ideal in $\mathfrak{E}$, the first assumption implies that $\mathfrak{c} = \mathfrak{E}$, and the two assumptions combined imply that the dual pairing $(\mathfrak{M}', \mathfrak{M}, \odot)$ is regular in the sense of §5. Since the ring $\overline{\mathfrak{B}}$ of all $\mathfrak{c}$-endomorphisms of $\mathfrak{M}$ is a centralizer of $\mathfrak{M}$ relative to $\mathfrak{c}$, the methods of §5 yield a correspondence between the $\overline{\mathfrak{B}}$-direct summands of $\mathfrak{M}$ and the left ideal direct components of $\mathfrak{c}$: to a $\overline{\mathfrak{B}}$-direct summand $\mathfrak{R}$ corresponds

$$\mathfrak{M}' \odot \mathfrak{R} = \{\psi_i \odot u_i \mid \psi_i \in \mathfrak{M}', \quad u_i \in \mathfrak{R}\},$$

while to a left ideal direct component $\mathfrak{l}$ of $\mathfrak{c}$ corresponds the $\overline{\mathfrak{B}}$-direct summand $\mathfrak{M}\mathfrak{l}$.

THEOREM 5. *Let $(\mathfrak{M}', \mathfrak{M}, \odot)$ be a regular pairing of $\mathfrak{M}' \times \mathfrak{M} \to \mathfrak{c}$, which is dual to an abstract pairing $(\mathfrak{M}', \mathfrak{M}, \tau)$. Then the mappings $\mathfrak{R} \to \mathfrak{M}' \odot \mathfrak{R}$ and $\mathfrak{l} \to \mathfrak{M}\mathfrak{l}$ between the set of $\mathfrak{B}$-direct summands of $\mathfrak{M}$ and the set of left ideal direct components of $\mathfrak{c}$ are inverses of each other. These mappings preserve direct sums and intersections whenever all modules concerned are direct summands. Two $\mathfrak{B}$-direct summands $\mathfrak{R}_1$ and $\mathfrak{R}_2$ are $\mathfrak{B}$-isomorphic if and only if $\mathfrak{M}' \odot \mathfrak{R}_1$ and $\mathfrak{M}' \odot \mathfrak{R}_2$ are $\mathfrak{c}$-isomorphic. $\mathfrak{R}$ is an indecomposable $\mathfrak{B}$-direct summand of $\mathfrak{M}$ if and only if $\mathfrak{M}' \odot \mathfrak{R}$ is an indecomposable left ideal in $\mathfrak{c}$.*

Proof. The first part of the theorem follows from Theorem 2, if we observe that a $\mathfrak{B}$-direct summand of $\mathfrak{M}$ is necessarily a $\overline{\mathfrak{B}}$-direct summand. By Theorem 2, a $\mathfrak{c}$-isomorphism between $\mathfrak{M}' \odot \mathfrak{R}_1$ and $\mathfrak{M}' \odot \mathfrak{R}_2$ induces a $\overline{\mathfrak{B}}$-isomorphism between $\mathfrak{R}_1$ and $\mathfrak{R}_2$, and hence $\mathfrak{R}_1$ and $\mathfrak{R}_2$ are $\mathfrak{B}$-isomorphic, since $\mathfrak{B}_r \subseteq \overline{\mathfrak{B}}$. Now let $x \to x^h$ be a $\mathfrak{B}$-isomorphism between $\mathfrak{R}_1$ and $\mathfrak{R}_2$. We supply the first step in the proof that

$$\sum \psi_i \odot x_i \to \sum \psi_i \odot x_i^h$$

is a $\mathfrak{c}$-isomorphism of $\mathfrak{M}' \odot \mathfrak{R}_1$ onto $\mathfrak{M}' \odot \mathfrak{R}_2$. Let $\sum \psi_i \odot x_i = 0$; then

$$\sum x_i \tau(\psi_i, \mathfrak{M}) = 0.$$

Since h is a $\mathfrak{B}$-isomorphism and $\tau(\psi_i, \mathfrak{M}) \subseteq \mathfrak{B}$, we have $\sum x_i^h \tau(\psi_i, \mathfrak{M}) = 0$. Then $\sum \psi_i \odot x_i^h = 0$, and the mapping is single valued. The rest of the proof is left to the reader. The final statement of the Theorem follows from the proof of the Corollary to Theorem 2.

Dually, we may state the following result.

THEOREM 5′. *Let $(\mathfrak{M}', \mathfrak{M}, \odot)$ be a regular pairing, as in Theorem 5. Then the mappings $\mathfrak{N}' \to \mathfrak{N}' \odot \mathfrak{M}$ and $\mathfrak{r} \to \mathfrak{r}\mathfrak{M}'$ between the $\mathfrak{B}$-direct summands of $\mathfrak{M}'$ and the right ideal direct components of $\mathfrak{c}$ have the properties stated in Theorem 5.*

We shall omit the proof of Theorem 5′.

8. Further results on the structure of $\mathfrak{c} = \mathfrak{M}' \odot \mathfrak{M}$. Using the results of §4, we shall establish a further theorem on the structure of the ring $\mathfrak{c} = \mathfrak{M}' \odot \mathfrak{M}$, in case the pairing $(\mathfrak{M}', \mathfrak{M}, \tau)$ is constructed from a projective representation of a finite group according to §2. In this case $\mathfrak{B} = \Delta(\mathfrak{G}, H, \rho)$ is a crossed product, and the set Ω is vacuous. We shall assume that the dual pairing $(\mathfrak{M}', \mathfrak{M}, \odot)$ is regular, so that the results of §7 are available.

THEOREM 7. *Let $(\mathfrak{M}', \mathfrak{M}, \tau)$ be a regular pairing of $\mathfrak{M}' \times \mathfrak{M} \to \mathfrak{B} = \Delta(\mathfrak{G}, H, \rho)$ as defined in §2. Let the dual pairing $(\mathfrak{M}', \mathfrak{M}, \odot)$ be regular. Then every indecomposable left or right ideal direct component of $\mathfrak{c} = \mathfrak{M}' \odot \mathfrak{M}$ contains a unique minimal subideal.*

Proof. First let $\mathfrak{l}$ be an indecomposable left ideal direct component of $\mathfrak{c}$. By Theorem 5, $\mathfrak{l} = \mathfrak{M}' \odot \mathfrak{N}$, where $\mathfrak{N}$ is an indecomposable $\mathfrak{B}$-direct summand of $\mathfrak{M}$. Our assumption that the pairing $(\mathfrak{M}', \mathfrak{M}, \odot)$ is regular implies that $\mathfrak{M}$ is a projective $\mathfrak{B}$-module, by Proposition 5 and the first remark thereafter. Therefore $\mathfrak{N}$ is $\mathfrak{B}$-isomorphic to an indecomposable right ideal direct component of $\mathfrak{B}$, and by Proposition 4, it follows that $\mathfrak{N}$ contains a unique minimal $\mathfrak{B}$-submodule $\mathfrak{m} \neq 0$. Since the pairing $(\mathfrak{M}', \mathfrak{M}, \odot)$ is non-degenerate, $\mathfrak{M}' \odot \mathfrak{m} \neq 0$. Now let $\mathfrak{l}' \neq 0$ be any left ideal contained in $\mathfrak{l}$. The fact that the pairing $(\mathfrak{M}', \mathfrak{M}, \tau)$ is regular implies that $\mathfrak{M}\mathfrak{l}' \neq 0$. By (20) we have

$$\mathfrak{l}' \supseteq \mathfrak{M}' \odot \mathfrak{M}\mathfrak{l}' \supseteq \mathfrak{M}' \odot \mathfrak{m},$$

and we have proved that $\mathfrak{M}' \odot \mathfrak{m}$ is the unique minimal subideal of $\mathfrak{l}$.

Now let $\mathfrak{r}$ be an indecomposable right ideal direct component of $\mathfrak{c}$. By Theorem 5′, $\mathfrak{r} = \mathfrak{N}' \odot \mathfrak{M}$, where $\mathfrak{N}'$ is an indecomposable $\mathfrak{B}$-direct summand of $\mathfrak{M}'$. By the second remark following Proposition 5, $\mathfrak{M}'$ is a projective $\mathfrak{B}$-module, and the argument given in the first part of the proof can be applied to prove that $\mathfrak{r}$ has a unique minimal subideal, as required.

COROLLARY. *Let Δ be a field, and let E be the subfield of Δ consisting of those elements of Δ left fixed by the automorphisms $\mathfrak{s}$, $s \in \mathfrak{G}$. Let the hypotheses of Theorem 6 be satisfied, and assume also that $\mathfrak{M}$ is finite dimensional over Δ. Then $\mathfrak{c} = \mathfrak{M}' \odot \mathfrak{M}$ is a QF-2 algebra*[6] *over the field* E.

[6] A finite-dimensional algebra $\mathfrak{A}$ over a field E is a QF-2 algebra (14) if every right or left ideal direct component of $\mathfrak{A}$ contains a unique minimal subideal.

Proof. It suffices to prove that $\mathfrak{c}$ is finite dimensional over E. Since Δ is commutative, the automorphisms $\bar{s}$, $s \in \mathfrak{G}$, form a finite group, and from Galois theory it follows that Δ is a finite extension of E. Therefore $\mathfrak{M}$ is finite dimensional over E. The elements of $\mathfrak{c}$ are l.t. in $\mathfrak{M}$ over E, and $\mathfrak{c}$ contains the scalar multiplications by elements of E, so that $\mathfrak{c}$ is a finite dimensional algebra over E, and the Corollary is proved.

Thrall's paper **(14)** contains a number of results concerning QF-2 algebras, all of which are directly applicable to $\mathfrak{c}$. We refer the reader to that paper for the details.

We add a final remark on the application of the theory to projective representations of groups. Let $(\mathfrak{M}', \mathfrak{M}, \tau)$ be a regular pairing of $\mathfrak{M}' \times \mathfrak{M} \to \mathfrak{B}$, constructed as in §2, and let $\mathfrak{C}$ be a centralizer of $\mathfrak{M}$ relative to $\mathfrak{B}$. Then Proposition 4, and the results of §5, can be applied to prove that every indecomposable $\mathfrak{C}$-direct summand of $\mathfrak{M}$ contains a unique minimal submodule. The proof is similar to the proof of Theorem 6, and will be omitted.

9. Applications to the Galois theory of primitive rings with minimal ideals. Let $\mathfrak{M}'$ and $\mathfrak{M}$ be left and right, respectively, vector spaces over a division ring Δ, which are dual relative to a non-degenerate bilinear form $\langle \psi, x \rangle$ on $\mathfrak{M}' \times \mathfrak{M} \to \Delta$. Let $\mathfrak{L}(\mathfrak{M}', \mathfrak{M})$ be the set of l.t. A on $\mathfrak{M}$ over Δ which possess transposes relative to the form $\langle \psi, x \rangle$, and let $\mathfrak{F}(\mathfrak{M}', \mathfrak{M})$ be the subset of $\mathfrak{L}(\mathfrak{M}', \mathfrak{M})$ consisting of finite valued l.t. We shall consider a ring $\mathfrak{A}$ of l.t. in $\mathfrak{M}$ over Δ such that **(7, 8)**

$$\mathfrak{F}(\mathfrak{M}', \mathfrak{M}) \subseteq \mathfrak{A} \subseteq \mathfrak{L}(\mathfrak{M}', \mathfrak{M}), \tag{25}$$

together with a finite group $\mathfrak{G}$ of automorphisms $A \to A^s$ of $\mathfrak{A}$. Then $\mathfrak{A}$ is a primitive ring with minimal ideals, and conversely, every primitive ring with minimal ideals is isomorphic to a dense ring of l.t. which satisfies (25). Let $\mathfrak{C}$ be the set of elements of $\mathfrak{A}$ which are left fixed by all the elements of $\mathfrak{G}$. We shall indicate how $\mathfrak{C}$ may be regarded as a centralizer of $\mathfrak{M}$ relative to a crossed product $\Delta(\mathfrak{G}, H, \rho)$, so that the results of §5–8 can be applied to discuss, for example, the subspaces of $\mathfrak{M}$ which are invariant relative to $\mathfrak{C}$.

For each element s in $\mathfrak{G}$, there exists a (1–1) s.l.t. U_s with associated automorphism $\bar{s}$ of $\mathfrak{M}$ onto itself, which possesses a transpose relative to the form, and which satisfies the equation

$$A^s = U_s^{-1} A U_s \tag{26}$$

for all $A \in \mathfrak{A}$. Since $(A^s)^t = A^{st}$, we obtain from (26),

$$U_t^{-1} U_s^{-1} A U_s U_t = U_{st}^{-1} A U_{st},$$

and

$$A U_s U_t U_{st}^{-1} = U_s U_t U_{st}^{-1} A,$$

for all A. Since $\mathfrak{A}$ is a dense ring of l.t., for each pair (s, t) there exists a scalar multiplication

$$\rho_{s,t}^{\overline{st}^{-1}}$$

such that

$$U_s U_t U_{st}^{-1} = \rho_{s,t}^{\overline{st}^{-1}},$$

or

$$(27) \qquad U_s U_t = U_{st} \rho_{s,t}.$$

It is now easy to verify that $\{\rho_{s,t}; \bar{s}\}$ is a factor set, and that if $\mathfrak{B} = \Delta(\mathfrak{G}, H, \rho)$ is the corresponding crossed product, then the mappings U_s define a representation of $\mathfrak{B}$ by endomorphisms of $\mathfrak{M}$. Since (26) is unchanged if we replace U_s by $U_s\mu_s$, we may assume that $\rho_{1,1} = 1$. Then the condition

$$\rho_{s,s^{-1}} = 1$$

of Proposition 1 is satisfied if and only if $U_s U_{s^{-1}} = U_1$ for all s in $\mathfrak{G}$. We have to show finally that $\mathfrak{C}$ satisfies (19). The elements of $\mathfrak{C}$ are $\mathfrak{B}$-endomorphisms of $\mathfrak{M}$. On the other hand, by (15) it follows that $\mathfrak{M}' \odot \mathfrak{M}$ is precisely the set of l.t. $\sum_s A^s$, where A ranges throughout $\mathfrak{F}(\mathfrak{M}', \mathfrak{M}) \subseteq \mathfrak{A}$, so that $\mathfrak{M}' \odot \mathfrak{M} \subseteq \mathfrak{C}$, and (19) is proved.

10. On the centralizer of a projective module. It seems probable that more penetrating results than we have obtained in §7 and 8 can be proved concerning the structure of the centralizer of a projective module. To support this view we shall prove the following result.

THEOREM 7. *Let $\mathfrak{A}$ be a commutative symmetric algebra of l.t. on a finite dimensional space $\mathfrak{M}$ over a field Φ such that $\mathfrak{M}$ is a unital projective (right) $\mathfrak{A}$-module. Then the centralizer $\mathfrak{C}$ of $\mathfrak{M}$ relative to $\mathfrak{A}$ is a symmetric algebra.*

Proof. The only consequences which we shall require of the assumption that $\mathfrak{M}$ is a projective $\mathfrak{A}$-module are the following: (a) the indecomposable direct summands of the $\mathfrak{A}$-module $\mathfrak{M}$ are $\mathfrak{A}$-isomorphic to indecomposable right ideal direct components of $\mathfrak{A}$ (**9**, Theorem 1); and (b) if $\mathfrak{M} = \mathfrak{M}_1 \oplus \mathfrak{M}_2$, where $\mathfrak{M}_1$ and $\mathfrak{M}_2$ are $\mathfrak{A}$-modules, then $\mathfrak{M}_1$ and $\mathfrak{M}_2$ are projective $\mathfrak{A}$-modules (**5**, p. 473).

We recall that $\mathfrak{A}$ is symmetric if and only if there exists a hyperplane $\mu(a) = 0$, which contains all commutators $ab - ba$ but no non-zero right or left ideals. We shall require the result that if $\mathfrak{A}$ and $\mathfrak{B}$ are symmetric algebras, then the Kronecker product $\mathfrak{A} \otimes \mathfrak{B}$ is symmetric.

Now we begin the proof of the theorem. First assume that $\mathfrak{A} = \mathfrak{A}_1 \oplus \mathfrak{A}_2$, where $\mathfrak{A}_1$ and $\mathfrak{A}_2$ are non-zero ideals. If we set $\mathfrak{M}_i = \mathfrak{M}\mathfrak{A}_i$ ($i = 1, 2$), then each $\mathfrak{M}_i$ is a faithful $\mathfrak{A}_i$-module, and $\mathfrak{M} = \mathfrak{M}_1 \oplus \mathfrak{M}_2$. The elements of the centralizer $\mathfrak{C}_i$ of $\mathfrak{M}_i$ relative to $\mathfrak{A}_i$ ($i = 1, 2$,) may be viewed as elements of the centralizer $\mathfrak{C}$ of $\mathfrak{M}$ relative to $\mathfrak{A}$, and with this agreement, $\mathfrak{C} = \mathfrak{C}_1 \oplus \mathfrak{C}_2$. It follows that $\mathfrak{C}$ is symmetric if we can prove that the $\mathfrak{C}_i$ are symmetric. Furthermore, each $\mathfrak{M}_i$ is a projective $\mathfrak{A}$-module, and hence a projective $\mathfrak{A}_i$-module ($i = 1, 2$). Thus we may assume, without loss of generality, that,

in addition to the hypotheses stated in the theorem, $\mathfrak{A}$ is an indecomposable algebra. Now let

$$\mathfrak{M} = \mathfrak{M}_1 \oplus \ldots \oplus \mathfrak{M}_s,$$

where the $\mathfrak{M}_i$ are indecomposable $\mathfrak{A}$-modules. Since $\mathfrak{M}$ is projective, each $\mathfrak{M}_i$ is $\mathfrak{A}$-isomorphic to $\mathfrak{A}$, by the indecomposability of $\mathfrak{A}$, and hence the $\mathfrak{M}_i$ are isomorphic to each other. Evidently $\mathfrak{M}_i$ is a faithful cyclic $\mathfrak{A}$-module. Since $\mathfrak{A}$ is commutative, the centralizer of $\mathfrak{M}_1$ relative to $\mathfrak{A}$ is isomorphic to $\mathfrak{A}$. The centralizer $\mathfrak{C}$ of $\mathfrak{M}$ is isomorphic to the full algebra of s by s matrices with coefficients in the centralizer of $\mathfrak{M}_1$ (**6**, p. 58), and hence

$$\mathfrak{C} \cong (\mathfrak{A})_s \cong \mathfrak{A} \otimes \Phi_s.$$

Since both $\mathfrak{A}$ and Φ_s are symmetric algebras, we conclude that $\mathfrak{C}$ is symmetric, and the theorem is proved.

11. Examples of regular pairings. We shall consider the pairing σ of §3, which has been studied by Weyl in connection with the representation theory of the full linear group. Let Φ be an arbitrary field of characteristic $p \geqslant 0$. Let $\mathfrak{M}$ be the m-fold Kronecker product with itself of an n-dimensional space $\mathfrak{B}$ over Φ. Let $\mathfrak{G} = \mathfrak{S}_m$ be the symmetric group on m letters, and let $b \rightarrow U(b)$ be the (ordinary) representation of the group algebra $\mathfrak{B}$ of $\mathfrak{G}$ by symmetry operators on $\mathfrak{M}$. Let $(\mathfrak{M}, \mathfrak{M}^*, \sigma)$ be the pairing defined in §3, and let $\mathfrak{b}$ be the nucleus $\sigma(\mathfrak{M}, \mathfrak{M}^*)$. We shall state without proof a few special results.

(a) $p > m$ or $p = 0$, n arbitrary. Then $\mathfrak{B}$ is semi-simple, and the pairing is regular. The centrally primitive idempotents of $\mathfrak{B}$ which are contained in $\mathfrak{b}$ have been determined explicitly by Weyl (**17**, Chap. IV).

(b) $m < n$, p arbitrary. Then $\mathfrak{b} = \mathfrak{B}$, and the pairing is regular.

(c) $m = 3$, $p = 3$, $n = 2$. Then $\mathfrak{b} = \mathfrak{B}$, and the pairing is regular. In this case the kernel $\mathfrak{K}$ of the representation U is different from zero, and $\mathfrak{b} \cap \mathfrak{K} = \mathfrak{K}$.

REFERENCES

1. E. Artin, C. J. Nesbitt, and R. M. Thrall, *Rings with minimum condition*, University of Michigan Publications in Mathematics, No. 1, 1944.
2. R. Brauer, *On sets of matrices with coefficients in a division ring*, Trans. Amer. Math. Soc., *49* (1951), 502–548.
3. H. Fitting, *Die Theorie der Automorphismenringe Abelscher Gruppen und ihr Analogen bei nicht kommutativen Gruppen*, Math. Ann., *107* (1932), 514–542.
4. W. Gaschütz, *Über der Fundamentalsatz von Maschke zur Darstellungstheorie der endlichen Gruppen*, Math. Z., *56* (1952), 376–387.
5. F. Kasch, *Grundlagen einer Theorie der Frobeniuserweiterungen*, Math. Ann., *127* (1954), 453–474.
6. N. Jacobson, *The theory of rings*, Mathematical Surveys, II (New York, 1943).
7. ———, *The radical and semi-simplicity for arbitrary rings*, Amer. J. Math., *67* (1945), 300–320.
8. ———, *Lectures in abstract algebra*, II (New York, 1953).

9. H. Nagao and T. Nakayama, *On the structure of* (M_0) *and* (M_u) *modules*, Math. Zeit., *59* (1953), 164–170.
10. T. Nakayama, *On Frobeniusean algebras*, I, Ann. Math., *40* (1939), 611–633.
11. ———, *On Frobeniusean algebras* II, Ann. Math., *42* (1941), 1–21.
12. C. Nesbitt, *On the regular representations of algebras*, Ann. Math., *39* (1938), 634–658.
13. C. Nesbitt and R. Thrall, *Some ring theorems with applications to modular representations*, Ann. Math., *47* (1946), 551–567.
14. R. Thrall, *Some generalizations of quasi-Frobenius algebras*, Trans. Amer. Math. Soc., *64* (1948), 173–183.
15. H. Weyl, *Commutator algebra of a finite group of collineations*, Duke Math. J., *3* (1937), 200–212.
16. ———, *The theory of groups and quantum mechanics* (New York, 1931).
17. ———, *The classical groups* (Princeton, 1939).

University of Wisconsin

THE COVERING OF SPACE BY SPHERES

E. S. BARNES

1. Introduction. Bambah (1) has recently determined the most economical covering of three dimensional space by equal spheres whose centres form a lattice, the density of this covering being

$$\vartheta_3 = \frac{5\sqrt{5}}{24}\pi. \tag{1.1}$$

As is well known, this problem may be interpreted in terms of the inhomogeneous minimum of a positive definite quadratic form. If $f(x) = f(x_1, x_2, \ldots, x_n)$ $(n \geq 2)$ is a positive quadratic form of determinant D, then, for any real $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$, we define $\mathfrak{m}(f; \alpha)$ to be the minimum of $f(x+\alpha)$ for integral x. The inhomogeneous minimum of $f(x)$ is then defined as

$$\mathfrak{m}(f) = \max_{\alpha} \mathfrak{m}(f; \alpha).$$

If now ϑ_n is the density of the most economical covering of n-dimensional space by lattice-ordered spheres, we have

$$\left(\frac{\vartheta_n}{J_n}\right)^{2/n} = \min_f \frac{\mathfrak{m}(f)}{D^{1/n}},$$

where J_n is the volume of the unit sphere:

$$x_1^2 + x_2^2 + \ldots + x_n^2 \leq 1.$$

Thus (1.1) is equivalent to the assertion that

$$\mathfrak{m}(f) \geq \left(\frac{125}{1024}D\right)^{\frac{1}{3}} \tag{1.2}$$

for all $f(x_1, x_2, x_3)$, and that the equality sign holds for some form f.

It is natural to introduce here the notion of an extreme form, by analogy with the corresponding homogeneous problem. We shall say that $f(x)$ is *extreme* if the ratio $\mathfrak{m}(f)/D^{1/n}$ is a (local) minimum, i.e. is not increased by any sufficiently small variation of the coefficients of f. Forms for which $\mathfrak{m}(f)/D^{1/n}$ is an absolute minimum may be called *absolutely extreme*. Since $\mathfrak{m}(f)$ and D are invariant under equivalence transformations (integral unimodular transformations of $x_1, \ldots, x_n$), while $\mathfrak{m}(f)/D^{1/n}$ is unaltered by multiplying f by an arbitrary positive constant, the property of being extreme is shared by the class of forms consisting of all forms equivalent to a multiple of some one form of the class.

Received September 13, 1955.

I prove here:

THEOREM 1. *If $n = 3$, there is just one class of extreme forms represented by*

$$(1.3) \qquad f_0(x_1, x_2, x_3) = 3x_1^2 + 3x_2^2 + 3x_3^2 - 2x_1x_2 - 2x_1x_3 - 2x_2x_3;$$

and for this class

$$(1.4) \qquad \mathfrak{m}(f) = \left(\frac{125}{1024} D\right)^{\frac{1}{3}}.$$

This theorem clearly includes the results of Bambah **(1)** (where the question of the existence of other classes of extreme forms is left open).

The object of this paper is, however, not so much to establish the above refinement of Bambah's results as to give a much simpler proof, which also suggests a method of attacking the problem when $n \geqslant 4$.

The starting point of the proof is Voronoï's method of reduction of a positive form f and the construction of the polyhedron Π associated with f. These are discussed in §2. Theorem 1 is proved in §3, while §4 contains some remarks on the method and the possibility of extending it to higher dimensions.

2. Reduced forms and their polyhedra. Voronoï **(3**, p. 150**)** has shown that every class of equivalent positive forms in 3 variables contains a form expressible as

$$(2.1) \quad f(x_1, x_2, x) = \rho_{01}x_1^2 + \rho_{02}x_2^2 + \rho_{03}x_3^2 + \rho_{12}(x_1 - x_2)^2 + \rho_{13}(x_1 - x_3)^2 + \rho_{23}(x_2 - x_3)^2$$

where

$$\rho_{ij} \geqslant 0 \qquad (i, j = 0, \ldots, 3);$$

and clearly the ρ_{ij} are uniquely determined by f. We call such a form *reduced* (in the sense of Voronoï).

The ρ_{ij} are not in general determined by the class of f. We have in fact, defining for convenience

$$\rho_{ij} = \rho_{ji}, \qquad i > j,$$

LEMMA 2.1. *If p, q, r, s is an arbitrary permutation of* 0, 1, 2, 3, *then the form*

$$(2.2) \quad \rho_{pq}x_1^2 + \rho_{pr}x_2^2 + \rho_{ps}x_3^2 + \rho_{qr}(x_1 - x_2)^2 + \rho_{qs}(x_1 - x_3)^2 + \rho_{rs}(x_2 - x_3)^2$$

is equivalent to the form (2.1).

Proof. The result is obvious if $p = 0$, since then (2.2) arises from (2.1) by the transformation $x_q \to x_1$, $x_r \to x_2$, $x_s \to x_3$. It therefore suffices to prove the result for $p, q, r, s = 1, 0, 2, 3$; this however corresponds to transforming (2.1) by

$$x_1 \to x_1, \quad x_2 \to x_1 - x_2, \quad x_3 \to x_1 - x_3.$$

This Lemma is the genesis of the suffix notation in (2.1), and provides an "argument by symmetry" which will be frequently used in what follows.

The set of points of space which are at least as near to the origin as to any integral point l (with the metric defined by f) forms a closed bounded convex polyhedron Π, the intersection of the half-spaces

$$f(x) \leqslant f(x - l),$$

where l runs through all integral points. Π may in fact be defined by a finite number $2\sigma \leqslant 2(2^3 - 1)$ of these inequalities of the type

$$f(x) \leqslant f(x \pm l_k) \qquad (k = 1, \ldots, \sigma).$$

The planes $f(x) = f(x \pm l_k)$ are then the faces of Π.

Perhaps the simplest method of obtaining $l_1, \ldots, l_\sigma$ is to use the criterion established by Voronoï (**4**, p. 277): a point $l(\neq 0)$ appears in the set $\pm l_1, \ldots, \pm l_\sigma$ if and only if the minimum of $f(x)$ over $x \equiv l \pmod 2$ is attained only for $x = \pm l$.

It is clear that, for the form (2.1), the minimum of $f(x)$ for prescribed parities of x_1, x_2, x_3 is attained when the even x_i are zero and the odd x_i are all 1 or all -1; and in general (e.g., if all $\rho_{ij} > 0$) only for these two sets. Thus, in general, Π has 7 pairs of parallel faces, for which we can find a symmetrical notation as follows:

Define $x_0 = 0$, so that

$$f = \sum_0^3 \rho_{ij}(x_i - x_j)^2,$$

and set

$$y_i = \tfrac{1}{2}\frac{\partial f}{\partial x_i} = \sum_{j=0}^3 \rho_{ij}(x_i - x_j) \qquad (i = 0, \ldots, 3);$$

then the 14 faces of Π are given by

$$\begin{aligned}
&F_i: && 2y_i = \sum_{l \neq i} \rho_{il},\\
(2.3)\qquad &F_{ij}: && 2(y_i + y_j) = \sum_{l \neq i,j} (\rho_{il} + \rho_{jl}),\\
&F_{ijk}: && 2(y_i + y_j + y_k) = \sum_{l \neq i,j,k} (\rho_{il} + \rho_{jl} + \rho_{kl}),
\end{aligned}$$

where all indices and summations run from 0 to 3. Since clearly $\sum y_i = 0$, the faces F_i, F_{jkl} and the faces F_{ij}, F_{kl} are parallel, where i, j, k, l is any permutation of 0, 1, 2, 3.

It is easy to verify the faces

$$\begin{aligned}
&F_1: 2y_1 = 2\rho_{01}x_1 + 2\rho_{12}(x_1 - x_2) + 2\rho_{13}(x_1 - x_3) = \rho_{01} + \rho_{12} + \rho_{13},\\
(2.4)\quad &F_{12}: 2(y_1 + y_2) = 2\rho_{01}x_1 + 2\rho_{02}x_2 + 2\rho_{13}(x_1 - x_3) + 2\rho_{23}(x_2 - x_3)\\
&\qquad\qquad = \rho_{01} + \rho_{02} + \rho_{13} + \rho_{23},\\
&F_{123}: 2(y_1 + y_2 + y_3) = 2\rho_{01}x_1 + 2\rho_{02}x_2 + 2\rho_{03}x_3 = \rho_{01} + \rho_{02} + \rho_{03},
\end{aligned}$$

determine a vertex v_{123} of Π; thus for example we have

$$|2(y_0 + y_1 + y_3)| = |2y_2| = |\rho_{02} - \rho_{12} + \rho_{23}| < \rho_{02} + \rho_{12} + \rho_{23}.$$

Applying all 4! permutations of the suffixes 0, 1, 2, 3, we obtain 4! distinct sets (F_i, F_{ij}, F_{ijk}) of faces determining 4! vertices v_{ijk}. Since Π has at most 4! vertices (**4**, p. 205), we have therefore determined all vertices of Π.

Our next task is to determine $\mathfrak{m}(f)$. From the definition of Π it is clear that

$$\mathfrak{m}(f) = \max_{x \in \Pi} f(x);$$

and by the convexity of Π and of the ellipsoid $f(x) < \mathfrak{m}(f)$, it follows that

$$\mathfrak{m}(f) = \max f(v)$$

over all vertices v of Π.

To calculate the values of $f(v)$, it suffices to evaluate $f(v_{123})$ and then to apply all permutations of suffixes in the ρ_{ij}; and the evaluation of $f(v_{123})$ may be simplified by observing that

$$f(x) = x_1y_1 + x_2y_2 + x_3y_3.$$

A direct calculation gives

$$4Df(v_{123}) = D(\rho_{01} + \rho_{02} + \rho_{03} + \rho_{12} + \rho_{13} + \rho_{23}) - K - 4\rho_{01}\rho_{03}\rho_{12}\rho_{23} \tag{2.5}$$

where D is the determinant of f (and of the equations (2.9)) and[1]

$$K = \sum \rho_{01}\rho_{02}\rho_{03}(\rho_{12} + \rho_{13} + \rho_{23}).$$

Since D, $\sum \rho_{ij}$ and K are invariant under permutation of suffixes of the ρ_{ij}, it follows from (2.5) that $f(v)$ has at most 3 distinct values for vertices v of Π. Denoting these by f_1, f_2, f_3 and setting

$$\lambda_1 = \rho_{01}\rho_{23},\ \lambda_2 = \rho_{02}\rho_{13},\ \lambda_3 = \rho_{03}\rho_{12}, \tag{2.6}$$

we have

$$\begin{aligned} 4Df_1 &= D(\textstyle\sum \rho_{ij}) - K - 4\lambda_2\lambda_3 \\ 4Df_2 &= D(\textstyle\sum \rho_{ij}) - K - 4\lambda_1\lambda_3 \\ 4Df_3 &= D(\textstyle\sum \rho_{ij}) - K - 4\lambda_1\lambda_2. \end{aligned} \tag{2.7}$$

Since $$D(f_i - f_j) = \lambda_k(\lambda_i - \lambda_j)$$

for i, j, k a permutation of 1, 2, 3, the value of

$$\mathfrak{m}(f) = \max(f_1, f_2, f_3) \tag{2.8}$$

is easily decided from the relative magnitudes of $\lambda_1, \lambda_2, \lambda_3$.

The above analysis has been carried out on the assumption that Π has 14

[1]We use here the usual summation convention for symmetric functions, so that K is the sum of the four distinct terms obtainable by cyclic permutations of 0, 1, 2, 3.

faces. If some of the ρ_{ij} vanish, some of the planes (2.4) are linearly dependent on the others and may be discarded. The effect of this is that certain of the 24 vertices coincide; thus if $\rho_{12} = \rho_{13} = \rho_{23} = 0$, Π degenerates to a parallelepiped, and $f(v)$ is the same for each of its 8 vertices. Such degeneration, however, does not affect the validity of our final results (2.7), (2.8).

It is convenient to note here, before proceeding to the proof of Theorem 1, some formulae concerning D, K and their derivatives.

We have

$$D = \begin{vmatrix} \rho_{01} + \rho_{12} + \rho_{13} & -\rho_{12} & -\rho_{13} \\ -\rho_{12} & \rho_{02} + \rho_{12} + \rho_{23} & -\rho_{23} \\ -\rho_{13} & -\rho_{23} & \rho_{03} + \rho_{13} + \rho_{23} \end{vmatrix}$$

$$(2.9) \qquad = \sum \rho_{01}\rho_{02}\rho_{03} + \sum \rho_{01}\rho_{23}(\rho_{02} + \rho_{03} + \rho_{12} + \rho_{13})$$

and, writing for convenience

$$(2.10) \qquad \sigma_i = \rho_{jk}\rho_{jl} + \rho_{jk}\rho_{kl} + \rho_{jl}\rho_{kl}$$

(where i, j, k, l is any permutation of 0, 1, 2, 3),

$$(2.11) \qquad \frac{\partial D}{\partial \rho_{01}} = \sigma_0 + \sigma_1 + \lambda_2 + \lambda_3.$$

Using symmetry, we obtain

$$(2.12) \qquad \frac{\partial D}{\partial \rho_{01}} - \frac{\partial D}{\partial \rho_{13}} = \sigma_0 - \sigma_3 - \lambda_1 + \lambda_2,$$

$$(2.13) \qquad \frac{\partial D}{\partial \rho_{01}} + \frac{\partial D}{\partial \rho_{23}} - \frac{\partial D}{\partial \rho_{02}} - \frac{\partial D}{\partial \rho_{13}} = -2(\lambda_1 - \lambda_2).$$

Similarly we find

$$(2.14) \qquad \frac{\partial K}{\partial \rho_{01}} = \rho_{02}\rho_{03}(\rho_{12} + \rho_{13} + \rho_{23}) + \rho_{12}\rho_{13}(\rho_{02} + \rho_{03} + \rho_{23}) + \rho_{02}\rho_{12}\rho_{23} + \rho_{03}\rho_{13}\rho_{23}$$

$$(2.15) \qquad \frac{\partial K}{\partial \rho_{01}} - \frac{\partial K}{\partial \rho_{13}} = (\lambda_2 - \lambda_1)(\rho_{03} + \rho_{12}) - \lambda_3(\rho_{01} - \rho_{02} + \rho_{23} - \rho_{13}) - (\rho_{03} + \rho_{12})(\rho_{01}\rho_{02} - \rho_{13}\rho_{23});$$

interchanging 1 and 2 and subtracting gives

$$(2.16) \qquad \frac{\partial K}{\partial \rho_{01}} + \frac{\partial K}{\partial \rho_{23}} - \frac{\partial K}{\partial \rho_{02}} - \frac{\partial K}{\partial \rho_{13}} = 2(\lambda_2 - \lambda_1)(\rho_{03} + \rho_{12}) - 2\lambda_3(\rho_{01} - \rho_{02} + \rho_{23} - \rho_{13}).$$

3. Proof of Theorem 1. We take f in the form (2.1), and suppose that f is extreme. We prove successively: (i) the two greater of $\lambda_1, \lambda_2, \lambda_3$ must be equal; (ii) $\lambda_1 = \lambda_2 = \lambda_3$; (iii) all ρ_{ij} are equal. In each case the proof proceeds by exhibiting a variation of the coefficients ρ_{ij} which, if the stated conditions

are not satisfied, contradicts our supposition that f is extreme. It will always suffice to work to the first order of small quantities; we denote generally by δR the first order variation in a function R of the ρ_{ij} resulting from small variations $\delta\rho_{ij}$.

In order to apply the analysis of §2 to both f and the neighbouring form $f' = \sum(\rho_{ij} + \delta\rho_{ij})(x_i - x_j)^2$, we must of course ensure that $\rho_{ij} + \delta\rho_{ij} \geqslant 0$ for all i, j. If all $\rho_{ij} > 0$, this will obviously hold for sufficiently small $\delta\rho_{ij}$ of either sign. If our hypotheses do not allow us to infer that $\delta\rho_{ij} \neq 0$ for some i, j we shall always choose the corresponding $\delta\rho_{ij} \geqslant 0$.

LEMMA 3.1. *If f is extreme, it is impossible that*

$$\lambda_1 > \lambda_2 \geqslant \lambda_3. \tag{3.1}$$

Proof. If (3.1) holds, we have $\mathfrak{m}(f) = f_1$, by (2.7), (2.8); and we shall have $\mathfrak{m}(f') = f'_1$ for any sufficiently near form f'.

We choose

$$\delta\rho_{01} = \delta\rho_{23} = -\epsilon, \ \delta\rho_{02} = \delta\rho_{13} = \epsilon \qquad (\epsilon > 0),$$

noting that (3.1) implies that $\rho_{01} > 0$, $\rho_{23} > 0$. Then, by (2.13),

$$\delta D = -\epsilon\left(\frac{\partial D}{\partial\rho_{01}} + \frac{\partial D}{\partial\rho_{23}} - \frac{\partial D}{\partial\rho_{02}} - \frac{\partial D}{\partial\rho_{13}}\right) = 2\epsilon(\lambda_1 - \lambda_2) > 0. \tag{3.2}$$

We set

$$L = D(\rho_{03} + \rho_{12}) - K - 4\lambda_2\lambda_3, \tag{3.3}$$

so that, by (2.7),

$$4f_1 = \rho_{01} + \rho_{02} + \rho_{13} + \rho_{23} + L/D. \tag{3.4}$$

Using (3.2) and (2.16) we find easily that

$$\begin{aligned}\delta L &= (\rho_{03} + \rho_{12})\,\delta D - \delta K - 4\delta(\lambda_2\lambda_3)\\ &= -2\epsilon\lambda_2(\rho_{01} + \rho_{02} + \rho_{23} + \rho_{13}),\end{aligned}$$

whence

$$\delta L < 0. \tag{3.5}$$

We have also

$$L \geqslant 0. \tag{3.6}$$

This may be verified by direct computation, using (3.3) and (3.1). We may argue more simply as follows:

Since $f(x) \geqslant f(1, 1, 0) = \rho_{01} + \rho_{02} + \rho_{13} + \rho_{23}$ for $x_1, x_2, x_3 \equiv 1, 1, 0 \pmod 2$, we have $f(x) \geqslant \frac{1}{4}(\rho_{01} + \rho_{02} + \rho_{13} + \rho_{23})$ for $x_1, x_2, x_3 \equiv \frac{1}{2}, \frac{1}{2}, 0 \pmod 1$; hence $\mathfrak{m}(f) \geqslant \frac{1}{4}(\rho_{01} + \rho_{02} + \rho_{13} + \rho_{23})$. Since $f_1 = \mathfrak{m}(f)$, (3.6) follows at once from (3.4).

We have thus shown that

$$\delta D > 0, \quad \delta f_1 < 0,$$

whence, for all sufficiently small $\epsilon > 0$,

$$\mathfrak{m}(f')D'^{-\frac{1}{4}} = f'_1 D'^{-\frac{1}{4}} < f_1 D^{-\frac{1}{4}} = \mathfrak{m}(f)D^{-\frac{1}{4}}.$$

This contradicts our assumption that f is extreme.

LEMMA 3.2. *If f is extreme, it is impossible that*

$$\lambda_1 = \lambda_2 > \lambda_3. \tag{3.7}$$

Proof. If (3.7) holds, we have $\mathfrak{m}(f) = f_1 = f_2 > f_3$; and, for any sufficiently near form f', $\mathfrak{m}(f') = \max (f'_1, f'_2)$. We choose

$$\delta\rho_{01} = \delta\rho_{23} = -\epsilon_1, \quad \delta\rho_{02} = \delta\rho_{13} = -\epsilon_2,$$
$$\delta\rho_{03} = \delta\rho_{12} = \epsilon_1 + \epsilon_2,$$

where $\epsilon_1 > 0$, $\epsilon_2 > 0$ (noting that (3.7) implies that $\rho_{01}, \rho_{23}, \rho_{02}, \rho_{13}$ are all positive). By restricting ϵ_1, ϵ_2 to satisfy

$$\epsilon_1(\rho_{01} + \rho_{23}) = \epsilon_2(\rho_{02} + \rho_{13}),$$

we ensure that

$$\delta(\lambda_1 - \lambda_2) = 0.$$

By (2.13) and (3.7), and writing for convenience

$$\lambda = \lambda_1 = \lambda_2,$$

we have

$$\begin{aligned}\delta D &= \epsilon_1\left(\frac{\partial D}{\partial\rho_{03}} + \frac{\partial D}{\partial\rho_{12}} - \frac{\partial D}{\partial\rho_{01}} - \frac{\partial D}{\partial\rho_{23}}\right) + \epsilon_2\left(\frac{\partial D}{\partial\rho_{03}} + \frac{\partial D}{\partial\rho_{12}} - \frac{\partial D}{\partial\rho_{02}} - \frac{\partial D}{\partial\rho_{13}}\right)\\ &= 2\epsilon_1(\lambda_1 - \lambda_3) + 2\epsilon_2(\lambda_2 - \lambda_3)\\ &= 2(\epsilon_1 + \epsilon_2)(\lambda - \lambda_3) > 0.\end{aligned}$$

We set

$$M = (\rho_{12} + \rho_{13} + \rho_{23})D - K - 4\lambda_2\lambda_3,$$

so that

$$4f_1 = (\rho_{01} + \rho_{02} + \rho_{03}) + M/D.$$

Arguing as in Lemma 3.1, we have

$$f_1 = \mathfrak{m}(f) \geqslant f(\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}) = \tfrac{1}{4}(\rho_{01} + \rho_{02} + \rho_{03}),$$

whence

$$M \geqslant 0.$$

Also, using (2.16) (with suitable permutations of the suffixes), we obtain

$$\begin{aligned}\delta M &= (\rho_{12} + \rho_{13} + \rho_{23})\delta D - \delta K - 4\delta(\lambda_2\lambda_3)\\ &= -(\epsilon_1 + \epsilon_2)[(\lambda - \lambda_3)(\rho_{01} + \rho_{02}) + \lambda\rho_{03} + \lambda_3\rho_{12}]\\ &< 0,\end{aligned}$$

since $\lambda > \lambda_3$, $\rho_{01} + \rho_{02} > 0$.

Since $\delta D > 0$, $\delta(\rho_{01} + \rho_{02} + \rho_{03}) = 0$ and $\delta M < 0$, we see that $\delta f_1 < 0$; and by symmetry $\delta f_2 < 0$. Hence for all sufficiently small ϵ_1, ϵ_2 we have

$$D' > D, \ \mathfrak{m}(f') < \mathfrak{m}(f),$$

contradicting our assumption that f is extreme.

LEMMA 3.3. *If f is extreme, then*

$$\lambda_1 = \lambda_2 = \lambda_3. \tag{3.8}$$

Proof. By a suitable permutation of suffixes we can ensure that $\lambda_1 \geqslant \lambda_2 \geqslant \lambda_3$; the result now follows from Lemmas 3.1, 3.2.

LEMMA 3.4. *If f is extreme, it is impossible that*

$$\rho_{01} > \rho_{13}, \ \rho_{02} > \rho_{23}. \tag{3.9}$$

Proof. Suppose that (3.9) holds. By Lemma 3.3, (3.8) holds and

$$\mathfrak{m}(f) = f_1 = f_2 = f_3.$$

We make the variation

$$\begin{aligned} -\delta\rho_{01} &= \delta\rho_{13} = \epsilon_1 = \epsilon(\rho_{01} + \rho_{13}), \\ -\delta\rho_{02} &= \delta\rho_{23} = \epsilon_2 = \epsilon(\rho_{02} + \rho_{23}), \\ \delta\rho_{03} &= -\epsilon_3 = -\epsilon(\rho_{01} + \rho_{02} - \rho_{13} - \rho_{23}), \\ \delta\rho_{12} &= 0, \end{aligned}$$

where $\epsilon > 0$. To justify this, we have to show that $\rho_{01} > 0$, $\rho_{02} > 0$, $\rho_{03} > 0$. Clearly $\rho_{01} > 0$, $\rho_{02} > 0$, by (3.9). If now $\rho_{03} = 0$, then $\lambda_3 = 0$, whence $\lambda_1 = \lambda_2 = 0$ by (3.8); this gives $\rho_{23} = 0$, $\rho_{13} = 0$, since $\rho_{01} \neq 0$, $\rho_{02} \neq 0$. But now $f = \rho_{01}x_1^2 + \rho_{02}x_2^2 + \rho_{12}(x_1 - x_2)^2$ and is clearly not positive definite.

It is easy to see that, for all sufficiently small ϵ, we have $\lambda'_1 = \lambda'_2 > \lambda'_3$, so that the neighbouring form f' has

$$\mathfrak{m}(f') = f'_1 \ (=f'_2).$$

For

$$\begin{aligned} \lambda'_1 - \lambda'_2 &= (\rho_{01} - \epsilon_1)(\rho_{23} + \epsilon_2) - (\rho_{02} - \epsilon_2)(\rho_{13} + \epsilon_1) \\ &= \lambda_1 - \lambda_2 - \epsilon_1(\rho_{02} + \rho_{23}) + \epsilon_2(\rho_{01} + \rho_{13}) = 0; \end{aligned}$$

and

$$\begin{aligned} \delta\lambda_1 &= \rho_{01}\epsilon_2 - \rho_{23}\epsilon_1 = \epsilon(\rho_{01}\rho_{02} - \rho_{13}\rho_{23}) > 0, \\ \delta\lambda_3 &= -\epsilon_3\rho_{12} \leqslant 0, \end{aligned}$$

so that $\lambda'_1 > \lambda'_3$.

We now obtain a contradiction to the fact that f is extreme by showing that

$$\delta D = 0, \quad \delta f_1 < 0. \tag{3.10}$$

By (2.12) and (2.11) we have

$$\begin{aligned}\frac{\partial D}{\partial \rho_{01}} - \frac{\partial D}{\partial \rho_{13}} &= \sigma_0 - \sigma_3 - \lambda_1 + \lambda_2 \\ &= (\rho_{02} + \rho_{12} + \rho_{23})(\rho_{13} + \rho_{23} - \rho_{01} - \rho_{02}) + \rho_{02}^2 - \rho_{23}^2,\end{aligned}$$

and, by symmetry,

$$\frac{\partial D}{\partial \rho_{02}} - \frac{\partial D}{\partial \rho_{23}} = (\rho_{01} + \rho_{12} + \rho_{13})(\rho_{13} + \rho_{23} - \rho_{01} - \rho_{02}) + \rho_{01}^2 - \rho_{13}^2.$$

Hence

$$\begin{aligned}
&- \epsilon_1\left(\frac{\partial D}{\partial \rho_{01}} - \frac{\partial D}{\partial \rho_{13}}\right) - \epsilon_2\left(\frac{\partial D}{\partial \rho_{02}} - \frac{\partial D}{\partial \rho_{23}}\right) \\
&= \epsilon(\rho_{01} + \rho_{02} - \rho_{13} - \rho_{23})[(\rho_{01} + \rho_{13})(\rho_{02} + \rho_{12} + \rho_{23}) \\
&\qquad\qquad + (\rho_{02} + \rho_{23})(\rho_{01} + \rho_{12} + \rho_{13})] \\
&\qquad - \epsilon(\rho_{01} + \rho_{13})(\rho_{02}^2 - \rho_{23}^2) - \epsilon(\rho_{02} + \rho_{23})(\rho_{01}^2 - \rho_{13}^2) \\
&= \epsilon_3[(\rho_{01} + \rho_{13})(\rho_{02} + \rho_{12} + \rho_{23}) + (\rho_{02} + \rho_{23})(\rho_{01} + \rho_{12} + \rho_{13}) \\
&\qquad - (\rho_{01} + \rho_{13})(\rho_{02} + \rho_{23})] \\
&= \epsilon_3(\sigma_0 + \sigma_3 + \lambda_1 + \lambda_3) \\
&= \epsilon_3 \frac{D}{\rho_{03}},
\end{aligned}$$

from which it follows immediately that $\delta D = 0$.

Writing, as in Lemma 3.1,

$$L = (\rho_{03} + \rho_{12})D - K - 4\lambda_2\lambda_3$$

we have, using $\delta D = 0$,

$$\delta L = D\delta\rho_{03} - \delta K - 4\delta(\lambda_2\lambda_3);$$

and a calculation similar to the above, using (2.14), (2.15) and (3.8), gives

$$\delta L = -2\epsilon_3\rho_{03}[(\rho_{01} + \rho_{12} + \rho_{13})(\rho_{02} + \rho_{12} + \rho_{23}) - \rho_{12}^2] < 0,$$

since $\epsilon_3 > 0$, $\rho_{03} > 0$. As in Lemma 3.1 we deduce that $\delta f_1 < 0$.

This establishes (3.10), and the Lemma is proved.

LEMMA 3.5. *If f is extreme, then*

$$\lambda_1 = \lambda_2 = \lambda_3 > 0. \tag{3.11}$$

Proof. By Lemma 3.3, it suffices to prove the impossibility of

$$\lambda_1 = \lambda_2 = \lambda_3 = 0. \tag{3.12}$$

Now if (3.12) holds, at least three ρ_{ij} are zero. Since in any three ρ_{ij} some suffix occurs at least twice, we may assume by symmetry that

$$\rho_{13} = \rho_{23} = 0. \tag{3.13}$$

Since $\lambda_3 = \rho_{03}\rho_{12} = 0$ and $\rho_{03} \neq 0$ (else f does not involve x_3 and so is not definite) we have $\rho_{12} = 0$. Thus

$$f = \rho_{01}x_1^2 + \rho_{02}x_2^2 + \rho_{03}x_3^2,$$

and, since f is definite, we have

$$\rho_{01} > 0, \quad \rho_{02} > 0. \tag{3.14}$$

Now (3.13) and (3.14) contradict Lemma 3.4.

LEMMA 3.6. *If f is extreme, then*

$$\rho_{01} = \rho_{02} = \rho_{03} = \rho_{12} = \rho_{13} = \rho_{23}. \tag{3.15}$$

Proof. We first show that $\rho_{01} = \rho_{13}$.

If $\rho_{01} \neq \rho_{13}$, then, after interchanging 0 and 3 if necessary, we have

$$\rho_{01} > \rho_{13}.$$

Since by Lemma 3.5

$$\lambda_1 = \rho_{01}\rho_{23} = \rho_{02}\rho_{13} = \lambda_2 > 0,$$

we have also

$$\rho_{02} > \rho_{23}.$$

By Lemma 3.4, these inequalities cannot hold.

Thus $\rho_{01} = \rho_{13}$. By symmetry we have

$$\rho_{ij} = \rho_{jk}$$

for any distinct suffixes i, j, k; from this (3.15) follows immediately.

Lemma 3.6 shows that the only possible class of extreme forms is that represented by

$$f_0(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 + (x_1 - x_2)^2 + (x_1 - x_3)^2 + (x_2 - x_3)^2;$$

and (1.4) of Theorem 1 is simply verified for $f = f_0$ by substituting $\rho_{ij} = 1$ in the formulae of §2.

Hence to complete the proof of Theorem 1 we have only to show that f_0 is in fact extreme. A direct proof of this is not difficult, but is rather tedious. It is simpler to appeal to a general theorem of Hlawka (2) which asserts the existence of a most economical lattice-covering of space, and hence the existence of a class of absolutely extreme forms (which can only be the class of f_0).

4. Remarks on the method. Voronoï (3; 4; 5) has given two distinct methods of reduction of positive quadratic forms. The first is based on the concept of perfect forms, and leads to a finite number of regions $R_0, R_1, \ldots, R_r$ in the $\frac{1}{2}n(n+1)$-dimensional coefficient space, with the properties: (i) any form is equivalent to a form lying in one of the regions R; (ii) no two forms lying in the interior of different regions are equivalent.

The second is based on the consideration of types of space-filling polytopes (which may be derived from positive forms, as we derived Π from f in §2), and leads to regions $R'_0, R'_1, \ldots, R'_r$ having the same two properties.

The "principal regions" R_0, R'_0 are derived respectively from the perfect form

$$\phi_0 = \sum_1^n x_i^2 + \sum_{i<j} x_i x_j$$

and its adjoint, a multiple of

$$f_0 = n \sum_1^n x_i^2 - 2 \sum_{i<j} x_i x_j;$$

and in fact $R_0 \equiv R'_0$.

For $n = 2$ and $n = 3$, $R_0 = R'_0$ is the only region, and we obtain for $n = 3$ the definition of reduction used in §2. For general $n > 2$, R_0 is the set of forms expressible as

$$(4.1) \qquad f(x) = \sum_{i<j} \rho_{ij}(x_i - x_j)^2, \; \rho_{ij} > 0 \; (i, j = 0, 1, \ldots, n), \; x_0 = 0.$$

It is to be noted that the regions R or R' do *not* possess the property that no two forms interior to the same region are equivalent; for example, the result of Lemma 2.1 generalizes in the obvious way for the form (4.1). This fact, which (as Voronoï remarks) is normally a disadvantage in a method of reduction, is clearly seen from the analysis of §§2 and 3 to be of considerable advantage in the problem we have been investigating. What Voronoï's second method of reduction achieves is the specification of the broadest type of forms whose polytopes Π (when not degenerate) are defined by the same set of integral points l; there is therefore little doubt that this method of reduction is best suited to the covering problem for each $n > 2$.

In conclusion, it is perhaps worth noting that the case $n = 2$ (for which there is just one region $R_0 = R'_0$) is very simply settled by these methods, and leads to

THEOREM 2. *If $n = 2$, there is just one class of extreme forms, represented by*

$$f_0(x_1, x_2) = x_1^2 + x_2^2 - x_1x_2,$$

and for this class

$$\mathrm{m}(f) = \left(\frac{4}{27} D\right)^{\frac{1}{2}}.$$

We take f in R_0, i.e.

$$f(x_1, x_2) = \rho_{01}x_1^2 + \rho_{02}x_2^2 + \rho_{12}(x_1 - x_2)^2, \quad \rho_{ij} > 0,$$

for which

$$D = \rho_{01}\rho_{02} + \rho_{01}\rho_{12} + \rho_{02}\rho_{12},$$

$$4D\mathrm{m}(f) = 4Df(v) = D(\rho_{01} + \rho_{02} + \rho_{12}) - \zeta_{01}\zeta_{02}\zeta_{12},$$

(the value of $f(v)$ being the same for all vertices v of Π).

If $\rho_{01} > \rho_{02}$, we take $\delta\rho_{01} = -\epsilon$, $\delta\rho_{02} = \epsilon$, $\epsilon > 0$, whence trivially D is increased and

$$4f(v) = (\rho_{01} + \rho_{02}) + \rho_{12}^2(\rho_{01} + \rho_{02})/D$$

is not increased; thus f cannot be extreme.

By symmetry it now follows that, for extreme f, we require $\rho_{01} = \rho_{02}$, and so $\rho_{01} = \rho_{02} = \rho_{12}$. Theorem 2 follows at once.

References

1. R. P. Bambah, *On lattice coverings by spheres*. Proc. Nat. Inst. Sci. India, *20* (1954), 25–52.
2. E. Hlawka, *Ausfüllung und Überdeckung konvexer Körper durch konvexe Körper*, Monatsh. Math. Phys., *53* (1949), 81–131.
3. G. Voronoï, *Sur quelques propriétés des formes quadratiques positives parfaites*, J. reine angew. Math. *133* (1907), 97–178.
4. ———, *Recherches sur les paralléloèdres primitifs* (Part 1), *ibid.*, *134* (1908), 198–287.
5. ——— (Part 2), *ibid.*, *136* (1909), 67–181.

University of Sydney

is

d

2.
h.

w.